

APLICACIÓN DE UN MODELO DE SOBREVIDA PARA EL ANÁLISIS DE LA
DESERCIÓN EN LOS PROGRAMAS DE ADMINISTRACIÓN DE LA
UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD
SECCIONAL DUITAMA



VÍCTOR RAÚL CAMARGO COLMENARES

Uptc

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA
SECCIONAL DUITAMA
ESCUELA DE MATEMÁTICAS Y ESTADÍSTICA
DUITAMA

2016

APLICACIÓN DE UN MODELO DE SOBREVIVENCIA PARA EL ANÁLISIS DE LA
DESERCIÓN EN LOS PROGRAMAS DE ADMINISTRACIÓN DE LA
UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD
SECCIONAL DUITAMA

VÍCTOR RAÚL CAMARGO COLMENARES
Código: 200921283

TRABAJO DE GRADO EN MODALIDAD MONOGRAFÍA:
para obtener el título de Licenciado en Matemáticas y Estadística.

DIRECTORA DEL PROYECTO:
CARMEN HELENA CEPEDA ARAQUE
MSc. Estadística.

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA
FACULTAD SECCIONAL DUITAMA
ESCUELA DE MATEMÁTICAS Y ESTADÍSTICA
DUITAMA
2016

Nota de Aceptación

Presidente del Jurado
MsC. CLARA EMILSE ROJAS MORALES

Jurado
Ms.C LEIDY ROCIO LEÓN DÁVILA

Jurado
Ms.C DAIRO SIGIFREDO GIL GIL

Duitama (1 de Abril 2016)

*Dedico este trabajo principalmente a
DÍOS, por haberme dado la vida y
permitirme haber llegado hasta este
momento tan importante a pesar de las
dificultades.*

AGRADECIMIENTOS

Agradezco a Dios por protegerme durante todo mi camino y darme fuerzas para superar obstáculos y dificultades en el tiempo de mis estudios.

A mi madre, por su ejemplo de mujer esforzada y valiente y por acompañarme durante todo el proceso de mis estudios.

A la directora del proyecto Ms.C Carmen Cepeda, por su apoyo siempre, por su dedicación, orientación, tiempo que pudo brindarme en cada reunión del proyecto, por su confianza y permitirme hacer parte de su equipo de trabajo.

A mis compañeros Yina Álvarez y Juan Carlos, por su compañía y apoyo en el proceso de la terminación del proyecto.

Gracias a todas las personas que ayudaron directa e indirectamente en la realización de este proyecto.

Victor R. Camargo

TÍTULO: APLICACIÓN DE UN MODELO DE SOBREVIDA PARA EL ANÁLISIS DE LA DESERCIÓN EN LOS PROGRAMAS DE ADMINISTRACIÓN DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA.

AUTOR: CAMARGO COLMENARES, Víctor Raúl.

DIRECTORA: CEPEDA ARAQUE, Carmen Helena.

PUBLICACIÓN: Duitama. Universidad Pedagógica y Tecnológica de Colombia, 2016.

UNIDAD PATROCINANTE: Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, Escuela de Matemáticas y Estadística.

PALABRAS CLAVES: Deserción estudiantil universitaria, modelos de sobrevida, Regresión de Cox.

OBJETIVO: Establecer a través de un modelo de sobrevida, las variables determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de los estudiantes para cada programa de Administración de la Facultad Seccional Duitama

DESCRIPCIÓN: este proyecto busca responder a la siguiente pregunta de investigación: ¿Cuáles son las variables que aparecen como fuertes determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de cada uno de los programas de Administración de la Facultad Seccional Duitama de la Universidad Pedagógica y Tecnológica de Colombia? El cuál se desarrolla a través de unos objetivos específicos, los cuales se mencionan como sigue:

Establecer el marco conceptual de la deserción estudiantil en la Facultad Seccional Duitama y, dependiendo de la disponibilidad de información, determinar las variables académicas y no académicas que se utilizarán para su medición en el tiempo.

Identificar el momento de inicio y la longitud del proceso de monitoreo de los estudiantes de los Programas Académicos, a partir de la información reportada en bases de datos de la Universidad.

Evaluar las variables académicas y no académicas relacionadas con la deserción mediante la construcción de un modelo de sobrevida que permita cuantificar la incidencia de cada una de ellas y determinar las más influyentes.

Usar la función de riesgo del modelo de sobrevida para cada uno de los estudiantes activos en I semestre de 2015 y calcular la probabilidad de abandono voluntario de sus estudios en el segundo semestre de 2015.

Diseñar un cuestionario de Excel para recoger información de variables no estáticas que permita a largo plazo construir un modelo de sobrevida.

FUENTES: Para el desarrollo de este proyecto se consultaron siete libros de estadística, específicamente de muestreo, análisis de datos con “*R*” y análisis de sobrevivencia, además se consultaron 7 artículos, en los cuales se describían estudios referentes la deserción y graduación estudiantil describiendo los factores que influyen sobre este. Para el caso de la obtención del marco muestral se acudió al Sistema de Información y Registro Académico de la Universidad Pedagógica y Tecnológica de Colombia. El diseño muestral correspondió a uno aleatorio simple sin reemplazo por cada programa académico.

CONTENIDO: En el documento se presenta siete partes, empezando por la presentación del proyecto, continuando con los referentes teóricos en el cual se muestra antecedentes investigativos sobre los factores que influyen en la deserción y uno relacionado con la modelación mediante el análisis de sobrevida. En seguida se presenta el diseño metodológico empleado. Una cuarta a la séptima parte, se construye para cada programa un capítulo con su respectivo modelo en deserción y graduación estudiantil, no sin antes presentar una descripción de los datos. Finalmente se dan unas conclusiones y recomendaciones.

METODOLOGÍA: Las fases que permitieron determinar la incidencia de los factores que aumentan el riesgo de la deserción de los programas de administración de la Universidad Pedagógica y Tecnológica de Colombia –Duitama, fueron:

1. Conformación del marco teórico el cual partió de la revisión documental sobre los antecedentes investigativos en la deserción estudiantil universitaria y, además lo referente a la teoría del análisis de Sobrevida.
2. Definición del diseño metodológico, el cual, entre otros aspectos, implica la el diseño muestral y definición de las variables explicativas obtenidas en el SIRA y SPADIES.
3. Conformación de las bases de datos por programa académico y cruce de la información, según fuentes seleccionadas.
4. Análisis, interpretación de la información y construcción del modelo de sobrevida que describe los factores que aumentan en el riesgo de desertar y graduarse en los estudiantes de administración de la UPTC Duitama.
5. Conclusiones y elaboración del informe final del proyecto.

CONCLUSIONES Y RECOMENDACIONES:

1. Al comparar la sobrevivencia por programa, se encontró que Administración Turística y Hotelera tiene una sobrevivencia superior (entendida la sobrevivencia como la continuación en el Programa académico) y los que tienen menos probabilidades de sobrevivir son los estudiantes de Administración Industrial, en cuanto a Administración de Empresas Agropecuarias se encuentra en el promedio de los dos programas anteriores, y de manera general las mujeres tienen mejor nivel de sobrevivencia respecto a los hombres.
2. En el programa de Administración de Empresas Agropecuarias, se tiene que la deserción es del 46%, cifra que según el reporte del MEN (2014) supera la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2%, sin embargo en cuanto a la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines es del 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 46, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría (46,45%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Pero también es considerable la proporción de aquellos que

lo hacen por causas no académicas predominando aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan de igual forma los hombres y mujeres, en el momento de presentar el ICFES no trabajaban, que además reportan no tener vivienda propia, cuya madre alcanza el nivel educativo de primaria, y sus ingresos familiares son de nivel bajo. El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, es decir del segundo al quinto semestre. Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Los factores que aumentan el riesgo de desertar son los de ser hombre y el de tener ingresos en nivel medio, hecho que es consistente con lo planteado por (Giovagnoli, 2002) quien lo atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa y al registrar ingresos familiares en nivel bajo, hecho particular que refleja que este tipo de estudiantes son mucho más comprometidos al hacer un esfuerzo mayor de avanzar dentro del Programa, posiblemente abarcando las oportunidades que ofrece la Universidad. Con base en lo anterior, la Universidad y el Programa en particular deberían diseñar acciones preventivas de acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración de Empresas agropecuarias es de 20 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas y tener un puntaje alto en el ICFES disminuyen la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gastar cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. Por el contrario registrar ingresos familiares en nivel bajo o medio, aumentan esta probabilidad de graduarse.

3. En cuanto al programa de Administración Industrial, se tiene que la deserción acumulada está cerca del 50%, cifra que supera según el reporte del MEN (2014)

la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 50, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría (59.6%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Sin embargo, también es considerable la proporción (40.4%) de los estudiantes que salen por causas académicas predominando aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan mayoritariamente los que en el momento de presentar el ICFES no trabajaban y que además reportan tener vivienda propia, cuya madre alcanza el nivel educativo de secundaria, y sus ingresos familiares son de nivel medio.

El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, entre el segundo y quinto semestre. Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Al observar los factores que explican el hecho de desertar se encontró que aumenta el riesgo de desertar en ser hombre, hecho que va de la mano con lo que sucede a nivel Nacional. Una explicación a este hecho es la dada por Giovagnoli, 2002 en donde se le atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa. Con base en ese hecho la Universidad y el Programa en particular deberían diseñar acciones preventivas de acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración Industrial es de 19 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas disminuye la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 13. Hecho que es preocupante ya que gasta tres

semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. En este punto valdría la pena hacer un análisis de los posgrados que ofrece la Facultad y que sean afines al área de administración.

4. En el programa de Administración turística y Hotelera se tiene que el 33,52% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 33, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Desertan mayoritariamente los hombres cuya edad promedio de ingreso es de 19 años, además de los estudiantes que ingresan al programa el 65 % tenían vivienda propia y el 84 % no trabajaban al momento de presentar el examen de estado, el puntaje promedio al momento de presentar el examen de estado fue de 67.6 también el 35,6 % de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de ATH, la posibilidad de que un estudiante se vaya en primer semestre es del 13%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Donde las mujeres tienen una sobrevivencia superior a los hombres en un 20%.

Los factores que aumentan el riesgo de desertar hacen referencia a ser hombre, hecho que corrobora (Giovagnoli, 2002) donde lo atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar con sus proyectos educativos que la mujer, y tener ingresos familiares de nivel bajo o medio, ya que posiblemente a los requerimientos del programa se deba cumplir con prácticas inter-

semestrales a diferentes lugares de la geografía colombiana o insumos de cocina o afines para la práctica de los conocimientos impartidos durante la carrera, se recomienda a la Universidad y el programa crear políticas de patrocinio y apoyo a aquellos estudiantes con este tipo de falacias económicas. Se disminuye el riesgo de desertar al presentar un alto puntaje en el ICFES en el momento de ingresar, entendiendo que la prueba es una buena medida de las habilidades y conocimientos, los cuales son fundamentales durante el avance del programa ya que éstos influyen en el resultado académico.

Por el contrario, para la graduación los factores que disminuyen esta propensión son una mayor edad de ingreso al programa y que registre tener un mayor número de asignaturas repetidas, posee menor probabilidad de graduarse. Además el semestre en el que es más probable que un estudiante se vaya es el primero y se graduó en el 14. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Se sugirieren actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica de asignación de temas de grado en semestres inferiores o vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

5. A partir de los resultados obtenidos se puede apreciar que los factores estudiados que afectan la deserción y la graduación en cada una de las Administraciones de la Facultad seccional Duitama (excepto el número de materias repetidas) son exógenos, difícilmente pueden ser cambiados por la institución a la hora de implementar estrategias y políticas para disminuirla. Los riesgos se asocian más con las características de la población estudiantil de la Facultad. De modo que, es importante que la UPTC encamine sus esfuerzos a mecanismos alternativos de estudio que les permita a estudiantes con ciertas características “desfavorables” reducir los riesgos de deserción y aumentar sus riesgos de graduación

6. El análisis de sobrevivencia es una potente herramienta, muy importante que ha venido cobrando fuerza no solo en el área médica, por la inclusión de la censura y otras características, sino en otros campos del saber en dónde permite a través del modelamiento establecer cuáles son los factores asociados a cierto fenómeno referidos al tiempo hasta la ocurrencia de un evento, en este caso permitió establecer los factores asociados a la deserción y graduación universitaria en algunos programas de la Facultad Seccional Duitama.

7. Se recomienda construir a futuro nuevos modelos de sobrevivencia a partir de la recolección de información para cada programa de administración de la Facultad Seccional Duitama, que contenga otros aspectos importantes y no medidos en este proyecto, para tal efecto este trabajo deja indicadas algunas variables y el formato de Excel para el registro, ver “Anexo H”.

8. Para estudios posteriores, se recomienda modelar el fenómeno de la deserción y graduación universitaria, de forma paramétrica asumiendo sobre el tiempo de deserción o graduación alguna distribución de probabilidad.

CONTENIDO

	Pág.
1. PRESENTACIÓN DEL PROYECTO	22
2. MARCO REFERENCIAL	28
2.1 ANTECEDENTES INVESTIGATIVOS	28
2.2. DESERCIÓN ESTUDIANTIL	33
2.3. TEORÍA DE MODELOS DE SOBREVIDA	36
2.3.1. Características de los Datos de Sobrevida	36
2.3.2. Especificación del Modelo de Sobrevida	40
2.3.1. El Estimador de Kaplan Meier	50
2.3.2. Estimador de la Función de Riesgo Acumulada.	51
2.3.3. Comparación de curvas de sobrevida	51
2.4. MODELO DE REGRESIÓN DE COX	53
2.4.1. Estimación de parámetros	54
2.4.2. Contraste de hipótesis del modelo de riesgos proporcionales	57
2.4.3. Método de selección del Modelo	60
2.4.4. Evaluación del modelo	62
2.4.5. Pronostico en el modelo de Riesgos Proporcionales	65
3. DISEÑO METODOLÓGICO	66
3.1. DISEÑO MUESTRAL	66
3.2. RECOLECCIÓN DE LA INFORMACIÓN	69
3.3. VARIABLES OBJETO DE ESTUDIO	70
3.4. HERRAMIENTAS PARA EL ANÁLISIS DE DATOS	72
4. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS	75
4.1. DESCRIPCIÓN DE LOS DATOS	75
4.2. MODELO DE SOBREVIDA PARA DESERCIÓN EN AEA	80
4.2.1 Función de Sobrevida	82

4.2.2. Factores asociados al riesgo de deserción	87
4.2.4. Pronóstico a partir del modelo	92
4.2.3. Evaluación del modelo	94
4.3. MODELO PARA GRADUACIÓN EN AEA	97
4.3.1. Función de Supervivencia.....	97
4.3.2 Factores relacionados al aumento de la graduación	98
4.3.3. Evaluación del Modelo.....	102
5. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN INDUSTRIAL	106
5.1. DESCRIPCIÓN DE LOS DATOS.....	106
5.2. MODELO DE SOBREVIDA PARA DESERCIÓN EN AI	111
5.2.1. Función de Supervida	112
5.2.2. Factores relacionados con el riesgo de Desertar.....	117
5.2.3. Evaluación del modelo de riesgos proporcionales.....	120
5.2.4. Pronóstico a partir del modelo	122
5.3. MODELO PARA LA GRADUACIÓN EN AI.....	124
5.3.1. Función de supervida.....	124
5.3.2. Factores asociados al aumento de graduación	126
5.3.3. Evaluación del Modelo.....	128
6. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN TURÍSTICA Y HOTELERA.....	132
6.1. DESCRIPCIÓN DE LOS DATOS.....	132
6.2. MODELO DE SOBREVIDA PARA LA DESERCIÓN PARA ATH.....	137
6.2.1. Función de Supervida	139
6.2.1. Factores asociados al riesgo de Deserción	142
6.2.3. Evaluación del modelo.....	147
6.2.4. Pronóstico a partir del modelo	149
6.3. MODELO PARA LA GRADUACIÓN DE ATH	150

6.3.1. Función de Sobrevida	151
6.3.2. Factores asociados al aumento de Graduación.....	152
6.3.3. Evaluación del modelo.....	155
CONCLUSIONES Y RECOMENDACIONES	159
BIBLIOGRAFÍA	166
ANEXOS	168

LISTA DE TABLAS

	Pág
TABLA 1 COMPARACIÓN DE CURVAS DE SOBREVIDA.....	52
TABLA 2 PARÁMETROS PARA DETERMINAR TAMAÑO DE MUESTRA.....	68
TABLA 3 VARIABLES OBJETO DE ESTUDIO.....	71
TABLA 4 RESUMEN DE CENSURA POR PROGRAMAS	74
TABLA 5 RESUMEN DESCRIPTIVO DE LAS VARIABLES DE ESTUDIO DEL PROGRAMA DE AEA	75
TABLA 6 CONTRASTES DE ALGUNAS VARIABLES CATEGÓRICAS AEA.....	78
TABLA 7 TEST DE CORRELACIÓN DE PEARSON.....	80
TABLA 8 CONSIDERACIONES PARA EL MODELO DE SOBREVIDA PARA AEA	80
TABLA 9 RESUMEN DE K-M PARA AEA.....	83
TABLA 10 ESTIMADOR K-M POR GÉNERO.....	85
TABLA 11 PRUEBAS DE VEROSIMILITUD, WALD Y SCORE.....	90
TABLA 12 FUNCIÓN DE SOBREVIDA DEL MODELO EN AEA.....	91
TABLA 13 PRONÓSTICO DEL MODELO INDIVIDUAL EN AEA	93
TABLA 14 ESTIMADOR DE K-M GRADUACIÓN EN AEA	97
TABLA 15 FUNCIÓN DE SOBREVIDA DEL MODELO PARA GRADUACIÓN.	101
TABLA 16 PRUEBAS DE VEROSIMILITUD, WALD Y SCORE.....	101
TABLA 17 RESUMEN DESCRIPTIVO DE LAS VARIABLES DE ESTUDIO DEL PROGRAMA DE AI.....	106
TABLA 18 CONTRASTES DE ALGUNAS VARIABLES CATEGÓRICAS DEL PROGRAMA DE AI	108
TABLA 19 TEST DE CORRELACIÓN DE PEARSON.....	110
TABLA 20 CONSIDERACIONES PARA EL MODELO DE SOBREVIDA PARA AI	111
TABLA 21 ESTIMADOR DE KM EN AI	113
TABLA 22 ESTIMADOR DE KM POR GÉNERO.....	115
TABLA 23. PRUEBAS DE VEROSIMILITUD, WALD Y SCORE.	119
TABLA 24 FUNCIÓN DE SOBREVIDA MODELO DE COX EN AI.....	119
TABLA 25 PRONÓSTICO DEL MODELO DE DESERCIÓN PARA AI.....	123
TABLA 26 ESTIMADOR DE KM PARA GRADUACIÓN.....	124
TABLA 27 PRUEBAS DE VEROSIMILITUD, WALD Y SCORE.....	128
TABLA 28 RESUMENES NÚMERICOS DE LA VARIABLES PARA ATH	132
TABLA 29 CONTRASTES DE ALGUNAS VARIABLES CATEGÓRICAS DEL PROGRAMA DE ATH.....	135
TABLA 30 TEST DE CORRELACIÓN DE PEARSON.....	137

TABLA 31 CONSIDERACIONES PARA EL MODELO DE SOBREVIDA PARA ATH	137
TABLA 32 ESTIMADOR DE KM EN ATH	139
TABLA 33 ESTIMADOR DE KM POR GÉNERO EN ATH.....	141
TABLA 34. PRUEBA DE VEROSIMILITUD, WALD Y SCORE.....	145
TABLA 35 FUNCIÓN DE SOBREVIDA MODELO DE DESERCIÓN EN ATH.....	145
TABLA 36 PRONÓSTICO DEL MODELO DE DESERCIÓN PARA AI.....	150
TABLA 37 MODELO PARA GRADUACIÓN EN ATH	154
TABLA 38 PRUEBAS DE VEROSIMILITUD, WALD Y SCORE.	155

LISTA DE GRÁFICAS

	Pág
GRÁFICA 1 (A) REPRESENTACIÓN DE ALGUNAS FUNCIONES DE RIESGO	46
GRÁFICA 2 FUNCIÓN DE SOBREVIVENCIA POR PROGRAMAS.....	73
GRÁFICA 3 FUNCIÓN DE SOBREVIVENCIA POR GÉNERO EN LAS ADMINISTRACIONES	74
GRÁFICA 4. DIAGRAMA DE LEXIS PARA AEA.....	81
GRÁFICA 5 FUNCIÓN DE SOBREVIVENCIA DE AEA	83
GRÁFICA 6 FUNCIÓN DE RIESGO ACUMULADA DE AEA	84
GRÁFICA 7 FUNCIÓN DE SOBREVIVENCIA POR GÉNERO DE AEA.....	85
GRÁFICA 8 OTRAS CURVAS DE SOBREVIVENCIA PARA AEA	86
GRÁFICA 9 FUNCIÓN DE SOBREVIVENCIA PARA EL MODELO DE AEA	91
GRÁFICA 10 FUNCIÓN DE RIESGO ACUMULADA DEL MODELO DE COX PARA AEA	92
GRÁFICA 11 RESIDUOS DE SCHOENFELD.....	95
GRÁFICA 12 RESIDUOS DE COX-SNELL PARA EL MODELO DE DESERCIÓN EN AEA	95
GRÁFICA 13 RESIDUOS DFBETA	96
GRÁFICA 14 FUNCIÓN DE SOBREVIVENCIA PARA LA GRADUACIÓN EN AEA.....	97
GRÁFICA 15 FUNCIÓN DE RIESGO ACUMULADA PARA GRADUACIÓN	98
GRÁFICA 16 FUNCIÓN DE SOBREVIVENCIA PARA EL MODELO DE GRADUACIÓN EN AEA.	100
GRÁFICA 17 FUNCIÓN DE RIESGO DEL MODELO DE GRADUACIÓN	101
GRÁFICA 18 RESIDUOS DE SCHOENFELD	102
GRÁFICA 19 RESIDUOS DE COX-SNELL PARA EL MODELO DE GRADUACIÓN EN AEA.....	103
GRÁFICA 20 RESIDUOS TIPO DFBETA	103
GRÁFICA 21 DIAGRAMA DE LEXIS PARA AI	112
GRÁFICA 22 FUNCIÓN DE SOBREVIVENCIA PARA AI	113
GRÁFICA 23 FUNCIÓN DE RIESGO ACUMULADA DE AI	114
GRÁFICA 24 FUNCIÓN DE SOBREVIVENCIA POR GÉNERO PARA AI	115
GRÁFICA 25 OTRAS CURVAS DE SOBREVIVENCIA PARA AI	116
GRÁFICA 26 FUNCIÓN DE SOBREVIVENCIA PARA MODELO DE COX PARA AI	119
GRÁFICA 27 FUNCIÓN DE RIESGO EN MODELO DE DESERCIÓN EN AI	120
GRÁFICA 28 RESIDUOS ESCALADOS DE SCHOENFELD	121
GRÁFICA 29 RESIDUOS DE COX-SNELL.....	121
GRÁFICA 30 RESIDUOS DF BETA	122
GRÁFICA 31 FUNCIÓN DE SOBREVIVENCIA PARA GRADUACIÓN EN AI	125

GRÁFICA 32 FUNCIÓN DE RIESGO ACUMULADA PARA LA GRADUACIÓN	125
GRÁFICA 33 FUNCIÓN DE SOBREVIVENCIA MODELO DE GRADUACIÓN EN AI	127
GRÁFICA 34 FUNCIÓN DE RIESGO MODELO DE GRADUACIÓN EN AI.....	127
GRÁFICA 35 RESIDUOS DE COX-SNELL PARA EL MODELO DE GRADUACIÓN EN AI	129
GRÁFICA 36 RESIDUOS DF BETAS	129
GRÁFICA 37 DIAGRAMA DE LEXIS – ATH	138
GRÁFICA 38 FUNCIÓN DE SOBREVIDA PARA ATH	139
GRÁFICA 39 FUNCIÓN DE SOBREVIDA POR GÉNERO PARA ATH	140
GRÁFICA 40 OTRAS CURVAS DE SOBREVIDA PARA ATH.....	141
GRÁFICA 41 FUNCIÓN DE SOBREVIDA MODELO DE DESERCIÓN EN ATH	146
GRÁFICA 42 FUNCIÓN DE RIESGO ACUMULADO MODELO DE DESERCIÓN DE ATH	146
GRÁFICA 43 RESIDUOS DE COX-SNELL.....	147
GRÁFICA 44 RESIDUOS ESCALADOS DE SCHOENFELD	148
GRÁFICA 45 FUNCIÓN DE SOBREVIDA DE GRADUACIÓN PARA ATH.....	151
GRÁFICA 46 FUNCIÓN DE RIESGO PARA GRADUACIÓN EN ATH	152
GRÁFICA 47 FUNCIÓN DE SOBREVIVENCIA DEL MODELO PARA GRADUACIÓN.....	154
GRÁFICA 48 FUNCIÓN DE RIESGO ACUMULADA MODELO DE GRADUACIÓN.....	154
GRÁFICA 49 RESIDUOS DE COX-SNELL.....	155
GRÁFICA 50 RESIDUOS, TIPO DF BETAS.....	156

LISTA DE ANEXOS

	Pág
ANEXO A. Selección de la Muestra para AEA.....	169
ANEXO B. Selección de la Muestra para AI.....	170
ANEXO C. Selección de la Muestra para ATH.....	171
ANEXO D. Metodo de selección del modelo de deserción por programa.....	173
ANEXO E. Sintaxis en R, Diagrama de Lexis	179
ANEXO F. Sintaxis en R por programas académicos.....	180
ANEXO G. Pronóstico para la cohorte I- 2015.....	191
ANEXO H. Formato de Registro de Información.....	194

1. PRESENTACIÓN DEL PROYECTO

Cifras como que el 45% de los estudiantes que ingresan a Educación Superior no completan sus estudios hasta graduarse (Educación de Calidad, el camino para la prosperidad, 2010-2014), justifican el desarrollo de proyectos como el de “Ampliación de la cobertura en la Educación Superior”, del Ministerio de Educación Nacional, que busca ampliar la cobertura apoyándose en acciones, entre otras, como la de promover en las Instituciones de Educación Superior (IES) acciones orientadas a fomentar la permanencia de los estudiantes en el sistema, aumentando su capacidad de monitorear e implementar estrategias para disminuir la deserción.

En el marco de este proyecto, constantemente el Ministerio de Educación Nacional -MEN- adelanta convocatorias como una estrategia para reducir la deserción estudiantil. A partir de la revisión de los objetivos de tales convocatorias, se puede afirmar que pese a que “la educación superior en los últimos años se han caracterizado por aumentos de cobertura e ingreso de estudiantes nuevos, el número de alumnos que logra culminar sus estudios es bajo, dejando entrever que una gran parte de estos los abandona, primordialmente en los primeros semestres” (Convocatoria para apoyar al fortalecimiento de proyectos institucionales para disminuir la deserción estudiantil en educación superior, link <http://www.mineducacion.gov.co/1621/article-133057.html>).

De hecho, la deserción estudiantil es una de las situaciones con las que se enfrentan cada una de las diferentes instituciones educativas en la sociedad, implicando tres problemas, el primero de ellos en el estudiante, ya que por los factores de superación y posicionamiento económico se genera frustración, el segundo es la universidad ya que pierde individuos y dinero porque invierte sus recursos en personas que abandonan sus proyectos, y por último la sociedad en donde se genera un incremento del subempleo.

Por ello el Ministerio de Educación Nacional ha liderado, en conjunto con las instituciones de educación superior, el diseño y la operación de una metodología de seguimiento de la deserción estudiantil en educación superior que se concreta en el Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES). Esta herramienta permite además de hacer un análisis descriptivo del riesgo de deserción, identificar y clasificar a los estudiantes en riesgo, dado un grupo determinado de variables. Estadísticamente hablando, lo que permite el SPADIES es estimar el riesgo de desertar en un punto determinado del tiempo (documento del MEN, página 40), nótese que hasta el año 2009, aproximadamente, el análisis de la deserción se había abordado desde un marco estático ignorando la evolución del fenómeno a lo largo del tiempo. Es decir, se describía por qué un estudiante decide abandonar sus estudios pero no se explicaba el proceso de abandono, específicamente, el tiempo que ha durado el estudiante hasta el momento de desertar. Por esa razón, en los últimos años el fenómeno de la deserción se ha venido estudiando como un proceso dinámico, a través de modelos como el de sobrevida, ya que permite determinar el riesgo de ocurrencia de un evento, en este caso el evento de desertar, y analizar cuándo es más probable que éste ocurra teniendo en cuenta la influencia de sus principales factores o predictores.

A partir de esta perspectiva el Ministerio de Educación Nacional con el fin de contribuir a que las instituciones de educación superior dispongan de información para el diseño de políticas “Antideserción” elaboró un estudio denominado “Deserción estudiantil en la educación superior Colombiana” con el propósito de contribuir a entender el problema de la deserción a través de su análisis en un modelo de sobrevida y como guía para el diagnóstico, diseño de acciones y evaluación de las mismas en las instituciones de educación superior. (Documento MEN, pág. 17)

Por supuesto el fenómeno de la deserción no es ajeno a la Universidad Pedagógica y Tecnológica de Colombia (UPTC) ubicada en el departamento de Boyacá. Institución que cuenta con cuatro sedes, 50 programas presenciales y 21 a distancia, de los cuales siete programas presenciales pertenecen a la seccional que se encuentra en la ciudad de Duitama, facultad interdisciplinar que cuenta con tres programas en administración: Turística y Hotelera, Empresas Agropecuarias e Industrial; dos programas de licenciatura uno en Tecnología y otro en Matemáticas y Estadística; uno de Ingeniería Electromecánica y uno en Diseño Industrial.

La Universidad, derivado del Plan de Desarrollo Institucional, Lineamiento Formación y Docencia, ha venido ejecutando el programa “Permanencia y Deserción”. Programa que busca caracterizar cualitativa y cuantitativamente la deserción para definir las causas de este fenómeno y establecer acciones para minimizarla tales como el Plan Padrino, el sistema de becas, monitores y los programas de bienestar. Con tales acciones se ha tratado de disminuir la deserción, tal como lo plantea el propósito del programa “Permanencia y Deserción”. Los estudios llevados a cabo, particularmente en la Facultad, han privilegiado el diagnóstico, el problema se ha abordado desde un marco estático ignorando la evolución del evento a lo largo del tiempo. Es decir, describen tan sólo cifras del comportamiento de la deserción y en algunos casos el por qué un estudiante decide abandonar sus estudios, aspecto importante pero no suficiente.

En ese sentido, este proyecto busca obtener una descripción y explicación de la naturaleza longitudinal del proceso de deserción, a través de un modelo de sobrevida, ya que dicha técnica de análisis permite seguir la variable dependiente hasta que ocurra el evento de interés, en nuestro caso la deserción. En consecuencia, este proyecto que se presenta busca responder a la siguiente pregunta de investigación:

¿Cuáles son las variables que aparecen como fuertes determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de cada uno de los programas de Administración de la Facultad Seccional Duitama de la Universidad Pedagógica y Tecnológica de Colombia?

Para abordar este interrogante de manera completa, el objetivo general del proyecto buscó: *“Establecer a través de un modelo de sobrevida, las variables determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de los estudiantes para cada programa de Administración de la Facultad Seccional Duitama”*. El cual se desarrolla a través de unos objetivos específicos, los cuales se mencionan como sigue:

Establecer el marco conceptual de la deserción estudiantil en la Facultad Seccional Duitama y, dependiendo de la disponibilidad de información, determinar las variables académicas y no académicas que se utilizarán para su medición en el tiempo.

Identificar el momento de inicio y la longitud del proceso de monitoreo de los estudiantes de los Programas Académicos, a partir de la información reportada en bases de datos de la Universidad.

Evaluar las variables académicas y no académicas relacionadas con la deserción mediante la construcción de un modelo de sobrevida que permita cuantificar la incidencia de cada una de ellas y determinar las más influyentes.

Usar la función de riesgo del modelo de sobrevida para cada uno de los estudiantes activos en I semestre de 2015 y calcular la probabilidad de abandono voluntario de sus estudios en el segundo semestre de 2015.

Diseñar un cuestionario de Excel para recoger información de variables no estáticas que permita a largo plazo construir un modelo de sobrevida que las

incluya tales como: relaciones intra-familiares y los problemas entre el estudiante y los miembros de su núcleo familiar; desintegración de los hogares por muerte de progenitores, separación o divorcio; migración al exterior, a otras regiones; el estado de salud, tanto física como mental del estudiante, enfermedades crónicas, incapacitantes, minusvalía; factores socioeconómicos; estudiantes con hijos, embarazos no planeados, matrimonio; situaciones del contexto político

Teniendo en cuenta las anteriores consideraciones, el desarrollo de este trabajo monográfico permitió, en primer lugar, al estudiante de Licenciatura en Matemáticas y Estadística poner en práctica su perfil ocupacional, educador y asesor estadístico en proyectos, ya que consiste en la aplicación de una técnica de modelamiento estadístico para analizar un tema relevante en educación, como lo es el de deserción. La técnica que se utilizó es un modelo de sobrevida, y resulta particularmente novedoso ya que a la fecha la deserción en los programas de la Facultad no ha sido abordada a través de ésta técnica estadística de dependencia.

En segundo lugar, el objeto de estudio que plantea esta monografía es el fenómeno de la deserción de los estudiantes, constituido en un problema sobre el cual debe actuar la Universidad, y que está directamente relacionado con la eficiencia de sus recursos y el cumplimiento de sus fines misionales. De hecho, el tema que se abordó resulta de gran interés para la comunidad académica, ya que mediante la interpretación de los parámetros estimados de un modelo de sobrevida, se proporciona a las autoridades académicas de la Facultad Seccional Duitama información sustentada y necesaria para formular posibles políticas o reformas educativas que aumenten la permanencia de los estudiantes dentro de la Institución.

Es claro que todos los estudiantes que abandonan una institución crean lugares vacantes que pudieron ser ocupados por alumnos que persistieran en los estudios. La pérdida de estudiantes causa problemas financieros a las instituciones al producir inestabilidad en la fuente de sus ingresos, pero también un alto costo social

que puede asociarse a la pérdida de productividad laboral derivada de la menor acumulación individual de capital humano.

En tercer lugar, tal como lo plantea el estudio denominado “Deserción estudiantil en la educación superior colombiana”, la deserción es un problema que afecta la relación del Estado con las instituciones de educación superior públicas, en el sentido del incumplimiento de las políticas y las metas sociales establecidas. Aspecto que resalta la relevancia de describir amplia y profundamente el comportamiento de la deserción en los programas de Administración de la Facultad.

En cuarto lugar, es evidente que el estudio de fenómenos como la deserción permite aportar elementos sustanciales en los procesos de autoevaluación, autorregulación y mejoramiento continuo de los Programas Académicos

2. MARCO REFERENCIAL

Diversos estudios se han realizado con el objetivo de determinar aquellas variables que influyen en la deserción estudiantil universitaria. Algunos de ellos estudios se describirán a continuación ya que sirven como antecedentes teóricos para la monografía que se presenta, además posteriormente se abordará lo relacionado con el fenómeno de la deserción y la teoría en modelos de sobrevivencia.

2.1 ANTECEDENTES INVESTIGATIVOS

Según (Osorio & Castillo, 2011) La mayoría de la literatura sobre la permanencia de los estudiantes en la universidad se enmarca en dos corrientes: el modelo de integración del estudiante (Tinto, 1975 y Spady, 1970) y el modelo de desgaste del estudiante (Bean, 1980); el primero hace referencia al proceso de integración y adaptación del estudiante a la vida universitaria, es decir, cuanto mayor sea el grado de integración al ambiente académico y social, mayor será su compromiso con la institución y por ende menor su probabilidad de desertar. El segundo hace referencia a variables externas a la institución que influyen en la probabilidad de desertar, atribuyendo mayor importancia relativa a los factores individuales.

Tinto (1989) afirma que el estudio de la deserción en la educación superior es extremadamente complejo, ya que implica no sólo una variedad de perspectivas sino también una gama de diferentes tipos de abandono. Adicionalmente, sostiene que ninguna definición puede captar en su totalidad la complejidad de este fenómeno, quedando en manos de los investigadores la elección de la aproximación que mejor se ajuste a sus objetivos y al problema por investigar.

A partir del modelo de integración del estudiante, algunas investigaciones utilizan datos longitudinales de cohortes de estudiantes, como el de Adelman (1999), que emplea datos del Centro Nacional de Estadísticas de Estados Unidos y plantea un

modelo de probabilidad lineal con estudiantes que ingresan a la universidad, para determinar los factores que explican el cambio en probabilidad de deserción; el autor encuentra que estos son los recursos académicos y la asistencia a clases. Por su parte, Robinson (1990) plantea la idea de que la permanencia en la universidad está directamente correlacionada con la relación del estudiante con sus profesores y compañeros, además, se observa que la probabilidad de desertar es mayor en el primer año académico. Basados en este tipo de modelo, Radcliffe, Huesman y Kellog (2006) plantean que las variables que mejor predicen la probabilidad y el momento de desertar son el rendimiento académico en el primer semestre y los antecedentes académicos de los estudios anteriores, así como la realización de cursos remediales de matemáticas y vivir fuera del campus universitario. En esta misma línea, Jones-White et (2008) encuentra que se presenta una desventaja competitiva cuando el estudiante no es admitido por primera vez en la universidad; de modo que el número de créditos aprobados, o los intentos por abandonar la universidad no son significativos para describir esta problemática.

Teniendo en cuenta el modelo de desgaste del estudiante, Willett y Singer(1991) plantean que la probabilidad de desertar es mayor para los estudiantes con menor rendimiento académico y que tienen padres con menor nivel educativo y menores ingresos. Este modelo también soporta el estudio de Cerioni y Ortíz de Guevara (1999), donde se concluye que los estudiantes de los últimos quintiles de ingresos tienen mayor probabilidad de permanencia en la universidad.

Montoya (1999) utiliza un modelo de duración y plantea que a mayores ingresos familiares mayor es el tiempo de permanencia del estudiante en la universidad, incluso más tiempo del considerado para graduarse.

Pagura et al. (2000), entre otras hipótesis, encuentran que el tiempo utilizado para concluir los dos primeros años de una carrera depende del género, del nivel de educación de los padres y de los ingresos. Concluyen que las características que favorecen terminar con éxito esos dos años son: el rendimiento académico, el

comienzo de la carrera justo después de concluir los estudios secundarios y el nivel educativo de los padres. En esta misma línea, Porto y Di Gresia (2004) plantean que los estudiantes con padres con nivel educativo bajo poseen una tasa de retención menor. Aina (2005) plantea que la baja educación de los padres y vivir solamente con uno de ellos aumenta la probabilidad de deserción. Además, pertenecer a una familia pequeña aumenta la probabilidad de graduarse. Por otra parte, la deserción es mayor en los primeros semestres, especialmente para los hombres y para los estudiantes que trabajan. El resultado más importante es que la decisión de retirarse de la universidad se ve claramente influenciada por las características de los padres, quienes al final son los responsables de la persistencia de los estudiantes.

Trabajos realizados en Panamá (2001), México (1989), Argentina y Costa Rica (1984), hacen referencia a aquellos estudiantes que realizan transferencias interinstitucionales que pueden entenderse como migrantes dentro del sistema, en los cuales las crisis económicas se convierten en un determinante de la deserción de instituciones privadas hacia instituciones públicas y del abandono definitivo del sistema educativo.

Por su parte, (Giovangnoli, 2002) utiliza un modelo de duración para identificar los factores que explican la deserción estudiantil y la graduación; dentro de sus principales conclusiones la autora plantea que el tipo de colegio, la educación de los padres, el género y la situación laboral del estudiante son los factores de mayor relevancia para determinar la probabilidad de que un estudiante abandone o no la universidad.

Por otra parte, existen estudios que combinan el modelo de integración y el de desgaste del estudiante. Por ejemplo, el Instituto Colombiano para la Evaluación de la Educación (ICFES) y la Universidad Nacional de Colombia (2002) encontraron que para Colombia, el género, la edad y las condiciones académicas y económicas son muy significativas, tanto para explicar la deserción como el rezago y la

graduación. Así mismo, las condiciones estructurales de la institución y la integración social, institucional y académica juegan un papel muy importante dentro de este análisis.

Stratton, O'toole y Wetzel (2008) encuentran diferencias entre los factores asociados a la deserción de corto plazo (interrupción de los estudios por menos de un año académico) y la deserción de largo plazo (interrupción de los estudios por más de un año académico). A través de un *logit* multinomial, los autores concluyen que estas diferencias se deben principalmente al momento en el que se matriculan los estudiantes, el tipo de ayuda financiera en el primer año y el estado civil. El estudio logra demostrar que gran parte de los abandonos son temporales y que los factores que afectan la deserción temporal no son los mismos que afectan la deserción definitiva.

Estudios sobre Deserción en Colombia

Algunas investigaciones relevantes realizadas por instituciones de educación superior o entes interesados como el MEN, se mencionan a continuación resaltando los logros obtenidos:

Cabrera *et al.* (1993) explican que debe realizarse una integración de los dos modelos (desgaste y de integración). Si se realiza este enlace entre ambas teorías, se puede comprender mejor la interacción y el papel que juegan las variables individuales, institucionales y ambientales en el momento de deserción de un estudiante. Así mismo, establecen que las estrategias dirigidas a evitar la deserción deben realizarse sobre aquellas variables que son predictoras y significativas, y sobre las cuales los administradores académicos pueden realizar alguna manipulación.

La Universidad de los Andes, a través del Centro de Estudios sobre Desarrollo Económico (CEDE) (2006), utiliza la información de las instituciones de educación

superior (IES) de Colombia hallando que algunos factores que aumentan el riesgo de deserción son: ser hombre, la educación de los padres, la ocupación laboral, el bajo puntaje en el examen de ingreso a la educación superior, estudiar programas como ingeniería, arquitectura y ciencias de la educación, estudiar en una universidad privada y no contar con apoyo financiero.

Castaño *et al.* (2006) plantean un análisis sobre los determinantes de la deserción y graduación en dos facultades de la Universidad de Antioquia (Colombia), realizado a partir de la aplicación de los modelos de riesgo proporcional en tiempo discreto de Prentice y Gloeckler (1978) y Meier (1990). Los resultados rechazan este último modelo y corroboran diferentes hipótesis sobre la utilización de cuatro conjuntos de factores: individuales, académicos, socioeconómicos e institucionales, como determinantes tanto de la deserción como de la graduación estudiantil. La hipótesis de que el estudiante bien informado tiene menor riesgo de desertar se corrobora.

En un estudio para el programa de Economía de la Universidad del Rosario (Colombia), Lopera (2008) encuentra, entre otros resultados, que la vinculación del estudiante con el mercado laboral y el hecho de que los estudiantes procedan de otras regiones aumentan el riesgo de deserción. Además, la edad del estudiante incrementa el riesgo y, sin embargo, su efecto decrece marginalmente.

El Ministerio de Educación Nacional de Colombia (2008), de acuerdo con la información del Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES), encuentra que la probabilidad de desertar aumenta según características: Individuales, socioeconómicas, académicas e Institucionales. En donde ser hombre, tener más de dos hermanos, menor educación de la madre, haber trabajado cuando presentó el examen de ingreso a la educación superior, enfrentar menores ingresos del hogar y obtener bajo puntaje en el examen del ICFES, aumentan el riesgo de desertar.

Osorio, A. y otros (2011). Utilizan dos modelos de riesgo proporcional para datos agrupados: el modelo de Prentice y Gloeckler (1978) y el modelo de Meier (1990). Los resultados sugieren que las características académicas más que las individuales son las que explican la decisión de abandonar los estudios. Específicamente, venir de otro programa académico y haber obtenido durante uno o varios semestres un promedio inferior a 3.25, afectan positivamente el riesgo de deserción. Por el contrario, para el caso de la graduación, ser hombre y pertenecer a Ciencias Económicas o Ingenierías, disminuyen la posibilidad de graduarse. Mientras que residir en la ciudad de Cali, tener mejores notas en matemáticas en el ICFES y un mejor rendimiento en el primer semestre, aumentan la probabilidad de graduarse.

2.2. DESERCIÓN ESTUDIANTIL

A pesar de las investigaciones acerca del tema aún se sigue discutiendo el concepto de deserción, hay consenso en definirla como un abandono que puede ser causado por factores socioeconómicos, individuales, institucionales y académicos. Concepto que varía desde el punto de vista que se aborde la investigación bien sea individual, estatal, institucional o nacional. Autores como Tinto (1989) afirman que por ser tan complejo el estudio de la deserción, ya que involucra gran variedad de variables, se deja al investigador que defina una aproximación al concepto de deserción, según los objetivos y el problema de la investigación. Aspecto que se establecerá a continuación.

Para la elaboración de esta monografía se tendrá en cuenta el reglamento estudiantil de la Universidad: Acuerdo 130 de 1998(A130/1998), así las cosas se considerará como desertor al estudiante que figure en el Sistema de Registro e Información Académico (SIRA) con: “PERDIO CUPO”A130/98” o “RETIRADO DEFINITIVAMENTE”. La categoría de “PERDIO CUPO” se refiere a los estudiantes que hayan perdido la calidad de este según lo contemplado en el Artículo 80 del

reglamento estudiantil y que se refiere a causas de tipo académico, a continuación se especifica cuáles son este tipo de causas:

“ARTÍCULO 80°. Pierde la calidad de estudiante por bajo rendimiento académico y no se le renovará la matrícula a quien se encuentre en una de las siguientes situaciones:

- a) Quien obtenga durante cuatro (4) semestres, un promedio aritmético acumulado inferior a tres cero (3.0).
- b) Quien teniendo un promedio aritmético acumulado inferior a tres cero (3.0), obtenga un promedio aritmético semestral inferior a dos cero (2.0).
- c) Quien pierda una asignatura que curse en calidad de repitente siendo su promedio aritmético acumulado inferior a tres cero (3.0). En el caso en que el promedio aritmético acumulado sea igual o superior a tres cero (3.0), la podrá cursar por tercera y última vez.
- d) Quien pierda en un mismo período académico dos asignaturas que se cursan en calidad de repitente.
- e) Quien pierda una asignatura que cursa por tercera vez”.

A lo anterior se suman los estudiantes que no renovaron matrícula dentro del plazo establecido en el Artículo 42 (tres semestres consecutivos). Se asumirá que la no renovación de matrícula tiene que ver con causas no académicas. La categoría de “Retiro Definitivo” se refiere al estudiante que le fue aprobada solicitud de retiro definitivo del programa y que, en teoría, se entenderá como causas no académicas.

La anterior definición de deserción se enmarca en lo que el MEN (2009) denomina el abandono de los estudios que puede ser forzoso o voluntario. Forzoso es cuando el sistema retira al estudiante por motivos de causa académica o disciplinaria. Cuando la iniciativa de deserción va por parte del estudiante se considera retiro voluntario.

Según el MEN en el año 2009, la deserción puede ser con respecto al tiempo y al espacio. Respecto al tiempo puede ser temporal o definitiva. La deserción temporal ocurre cuando el estudiante ha reservado cupo y se espera su reintegro en determinado tiempo o cuando reingresa después de haber sido sancionado académicamente o haberse retirado. Estas categorías se pueden subdividir a la vez en:

Deserción precoz: individuo que habiendo sido admitido por la institución de educación superior no se matricula. Para efectos del estudio se asumirá como el estudiante que cursa menos de un semestre.

Deserción temprana: individuo que abandona sus estudios en los primeros semestres del programa. Para efectos del estudio corresponde antes o habiendo cursado el 50% del programa. [2 - 5 semestre].

Deserción tardía: individuo que abandona los estudios en los últimos semestres. Para efectos del estudio corresponde después de haber cursado el 50% del programa.[6 semestre- en adelante]

La deserción con respecto al espacio, por su parte, se divide en:

Deserción institucional: caso en el cual el estudiante abandona la institución.

Deserción interna o del programa académico: se refiere al alumno que decide cambiarse a otro programa que ofrece la misma institución de educación superior a este tipo de deserción se le conoce también como *trasferencias o movilidad estudiantil*.

De acuerdo a los tipos de deserción descritos se establece que para los propósitos de este proyecto sólo se contemplará la deserción con respecto al tiempo – definitiva, no se tendrá en cuenta el espacio ya que se tiene únicamente la trazabilidad del estudiante en los programas objeto de estudio.

2.3. TEORÍA DE MODELOS DE SOBREVIDA

En estudios internacionales sobre la deserción estudiantil, se han aplicado diferentes metodologías con el objeto de estimar el riesgo de desertar en un punto determinado del tiempo. Se han usado los modelos de variables dependientes discretas o cualitativas, en particular, los modelos de regresión logit, probit además el análisis discriminante. Sin embargo el análisis de supervivencia ha venido tomando gran importancia, para el estudio de la deserción estudiantil. La razón de este crecimiento es el que permite, a diferencia de los otros modelos, ya mencionados, un análisis más dinámico del fenómeno, es decir permiten hacer un seguimiento de los estudiantes desde que inician la carrera hasta que se presente alguno de los eventos posibles (en nuestro caso la deserción), y relacionarlos con un conjunto de variables (individuales, socioeconómicas, institucionales y académicas).

El análisis de supervivencia originalmente se aplicó en Medicina con tablas de mortalidad, de ahí proviene la denominación y terminología, este utiliza cuando la variable dependiente representa un periodo de tiempo entre un suceso inicial y el evento final. El análisis de Supervivencia (Survival Analysis) tiene por objeto conocer el efecto de una variable independiente cuando la variable dependiente puede expresarse en términos de tiempo hasta que se da determinado suceso. Es decir, dada una variable cuyos valores corresponden al tiempo que transcurre hasta que ocurre un determinado suceso final, el objetivo del análisis es estimar, en función del tiempo, la probabilidad de que ocurra dicho suceso.

2.3.1. Características de los Datos de Sobrevida

Según (Godoy, 2009) en el análisis de sobrevida el interés se centra en un grupo o varios, con determinadas características, a los cuales, se les observará un evento

determinado, llamado falla, el cual ocurre dentro de un lapso de tiempo, llamado tiempo de falla, dicha falla solo puede ocurrir máximo una vez en cualquier individuo. El tiempo de supervivencia es el que ocurre desde el momento que entra al estudio o tiempo inicial hasta el tiempo final o tiempo que transcurre hasta la ocurrencia del evento de interés.

(Díaz, 2015) Menciona una estructura para las características de los datos de sobrevivencia, así:

- La unidad sobre la cual se registra el evento.
- El evento de interés.
- El tiempo origen o inicial y el tiempo final del estudio.
- La escala de medida del tiempo hasta el evento.
- El tipo de censura y el truncamiento del tiempo.

Tiempo de Falla

Según (Colosimo & Ruiz Giolo, 2006) es el tiempo de sobrevivencia o tiempo de vida, denotado por T , donde esta es una variable aleatoria real no negativa. En campo de algunas aplicaciones, el valor de T es el tiempo transcurrido hasta la ocurrencia de un evento (falla). Por ejemplo:

En un ensayo clínico, T es el tiempo desde el inicio del tratamiento de un paciente hasta su muerte.

En el estudio de enfermedades infecciosas, T representa el tiempo desde el ataque de la infección hasta la aparición de los síntomas.

En el estudio de las enfermedades genéticas, T representa desde el nacimiento hasta el inicio de la enfermedad genética, en este caso, T es la edad del individuo.

Para el caso de la monografía que se presenta, el tiempo de falla T corresponde al tiempo en semestres en el que un estudiante deserta de alguno de los programas objeto de estudio.

(Borges, 2005) menciona dos mecanismos que no hacen posible la observación completa ya que dificultan el análisis de los tiempos de seguimiento, según como lo son la censura y el Truncamiento. En cuanto a la censura existen dos tipos: censura tipo I en la cual los individuos son observados hasta un tiempo determinado y, la censura tipo II en la cual los individuos son observados hasta que ocurran un número determinado de fallas o eventos de interés.

Los mecanismos de censura (Tipo I) y truncamiento más frecuentes son presentados a continuación como lo describe (Godoy, 2009):

Censura

Esta ocurre cuando se sabe que algunos tiempos de falla han ocurrido solamente dentro de ciertos intervalos y el resto de los tiempos de vida son conocidos exactamente. Hay varias categorías de censura, principalmente Censura por la derecha (Tipo I, Tipo II y Tipo III), Censura por la izquierda y Censura por intervalo.

a. *Censura Tipo I (a derecha)*: Se presenta cuando hasta la última observación que se le hace al individuo, aún no se ha ocurrido el evento que se desea observar. Se considera una población homogénea de individuos (para el caso Estudiantes Universitarios), $i = 1, \dots, n$, los cuales pueden experimentar el evento de interés (para el caso deserción) sin reemplazo, es decir que una vez ocurra el evento no es posible que vuelva a ocurrir (lo que significa que no se tendrá en cuenta en el estudio los casos de amnistía). Además dicha población es observada por un periodo de tiempo limitado a partir de $t = 0$, en el cual el individuo puede presentar o no el evento de interés, obteniendo así información incompleta en este caso, sobre la ocurrencia del evento, por lo que se considera al individuo censurado.

Existen varias razones según (Borges, 2005) para que se presente este tipo de censura:

- Que hasta el momento de la finalización del estudio no haya ocurrido el evento, esto ocurrirá en el caso de que el período de seguimiento sea finito. (Para el caso de este proyecto se presenta censura cuando el Estado del estudiante en el SIRA está registra de las siguientes formas: Graduado, Matriculado-Terminación Académica).
- Que el individuo haya abandonado el estudio.
- Que haya ocurrido en el individuo otro evento que imposibilite la ocurrencia del evento que se desea observar.

b. Censura a derecha tipo II: Un segundo tipo de censura por la derecha es la Censura tipo II, en la cual hay dependencia del tamaño de muestra (denotado por n) y las fallas que se observen. En esta, todos los individuos son puestos en estudio al mismo tiempo y se da el término de este cuando r de los n individuos han presentado el evento de interés. En este caso, el tiempo de censura es aleatorio, pues $(n - r)$ observaciones serán censuradas al tiempo dado por la r -ésima falla, la cual no se sabe cuándo ocurrirá. De modo que esto marca una diferencia importante entre la Censura de Tipo I y la Censura tipo II(r).

c. Censura por la izquierda: Es poco común en análisis de supervivencia, se presenta cuando para la primera observación que se realiza sobre el individuo ya ha ocurrido el evento que se desea observar.

d. Censura por intervalos: Se presenta cuando solo se sabe que al individuo le ocurre el evento de interés entre un instante t y un tiempo t .

Nótese que para este proyecto solo se abordará la censura a derecha tipo I.

Truncamiento

(Godoy, 2009) define el truncamiento como una condición que presentan ciertos sujetos en el estudio y el investigador no puede considerar su existencia. Cuando los datos presentan truncamiento, solamente individuos a los que les ocurre algún evento particular, antes del evento de interés o la censura, son considerados en el análisis por el investigador. Dentro de este se encuentra el truncamiento izquierda y a derecha, como se describen a continuación:

a. Entrada tardía al estudio (truncamiento por la izquierda): Este ocurre cuando los sujetos entran al estudio a una edad particular (no necesariamente el origen del evento de interés), y son observados desde este “tiempo retrasado de entrada”, hasta que el evento ocurra o hasta que el evento es censurado.

b. Truncamiento por la derecha: Este ocurre cuando solo individuos que han presentado el evento son incluidos en la muestra y ningún sujeto que no haya presentado aún el evento será considerado. Un ejemplo de muestras que presentan truncamiento por la derecha, son los estudios de mortalidad basados en registros de muerte.

2.3.2. Especificación del Modelo de Sobrevida

(Godoy, 2009) menciona que un modelo de sobrevivencia es caracterizado por variables aleatorias no negativas, así La variable aleatoria no negativa T que representa el tiempo corresponde a $T \geq 0$, la cual denota los tiempos de duración de los individuos de la población siendo está monitoreada en el tiempo hasta que se presente o no el evento, la variable de interés en los modelos de sobrevida es el tiempo de duración hasta la ocurrencia de éste, es decir, para el caso del proyecto corresponde al tiempo que ha durado el estudiante hasta el momento de desertar.

(Diaz, 2015) muestra que esta variable con función de densidad de probabilidad $f(t)$ y función de distribución acumulada $F(t) = P(T \leq t)$, se expresa como:

$$P(T \leq t) = \begin{cases} \sum_{t_j \leq t} p(t_j) = F(t), & \text{caso discreto.} \\ \int_0^t f(u)du, & \text{caso continuo.} \end{cases}$$

La distribución de esta variable puede ser caracterizada por al menos 4 funciones básicas: función de supervivencia, función de riesgo, función de densidad de probabilidad y la función de Vida media residual. Estas funciones son matemáticamente equivalentes, en el sentido de que si una de ellas está dada, pueden derivarse las otras tres.

2.3.2.1. Función de Sobrevida

(Godoy, 2009) menciona que la función básica empleada para describir los fenómenos de tiempo-evento es la función de sobrevida denotada por $S(t)$, también llamada *tasa de supervivencia acumulativa*. Esta función es la probabilidad de que un sujeto en estudio no experimente el evento de interés (sobreviva) antes de un momento dado, por tanto, sea T , una variable aleatoria no negativa (de tiempo de falla) con función de distribución $F(t)$ y función de densidad de probabilidad $f(t)$, a continuación se describe estas posibilidades:

Función de distribución acumulada: $F(t) = P(T \leq t)$, representa la probabilidad de que el evento de interés ocurra en un tiempo menor o igual que t .

Para el análisis de sobrevivencia, es de especial interés la función de sobrevida:

$$\begin{aligned} S(t) &= P(T > t) \\ &= P(\text{un individuo sobreviva mas allá de } t). \end{aligned}$$

O visto de otra forma:

$$\begin{aligned}
S(t) &= 1 - F(t) \\
&= 1 - P(T \leq t) \\
&= 1 - P(\text{un individuo falle antes del tiempo } t).
\end{aligned}$$

Por tales características, $S(t)$ es una función no creciente tal que:

$$S(0) = 1 \quad \text{y} \quad S(t) = 0 \quad \text{Cuando } (t \rightarrow \infty).$$

2.3.2.2. Función de Densidad de probabilidad

Como cualquier variable aleatoria, el tiempo de supervivencia T tiene una función de densidad de probabilidad. En el caso continuo:

$$S(t) = P(T > t) = \int_t^{\infty} f(u) du = 1 - F(t)$$

Entonces, $f(t) = -\frac{dS(t)}{dt}$ de esta relación $f(t)dt$ puede ser de manera aproximada la probabilidad de que un evento pueda ocurrir al tiempo t y $f(t)$ es una función no negativa con área bajo $f(t)$ igual a uno.

Para el caso discreto, como lo menciona (Diaz, 2015) se supone que T toma valores t_j , con $j=1,2,\dots$, con función de probabilidad $p(t_j) = P(T = t_j)$, $j= 1,2,\dots$, donde se asume que $t_1 < t_2 \dots$

La función de supervivencia para la variable aleatoria discreta T está dada por:

$$S(t) = P(T > t_j) = \sum_{t_j > t} p(t_j).$$

Un ejemplo de la función de densidad, puede ser encontrada la proporción de individuos que cae en cualquier intervalo de tiempo y el pico de frecuencia más alto

de fallas. A continuación se muestra un ejemplo en la figura 1 de una curva de supervivencia sin evento trombótico en 26 pacientes que fueron seguidos a largo plazo después del alta hospitalaria. Casi un 30% de los pacientes se vieron afectados por un nuevo accidente trombo embólico en los 2 años que siguieron al acontecimiento índice.

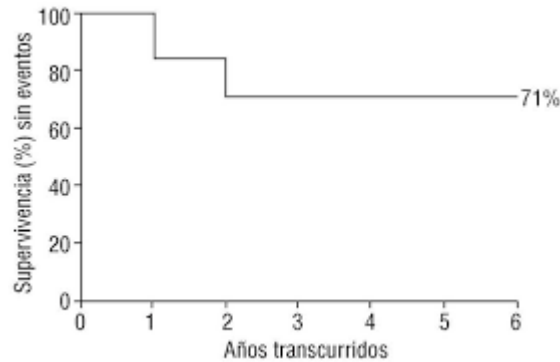


Figura 1 Curva de Densidad de una Función de Supervivencia.

2.2.2.4. Función de Riesgo

Otra función de interés para el análisis es la llamada **función de riesgo** $h(t)$ o (λ) que representa el riesgo instantáneo de que un evento ocurra en un intervalo $[t_1, t_2)$ puede ser expresada en términos de una función de supervivencia como:

$$S(t_1) - S(t_2)$$

La tasa de falla para el intervalo $[t_1, t_2)$ está definida como la probabilidad de que la falla ocurra en ese intervalo, dado que no ocurre antes de t_1 , así esta tasa de falla o riesgo en el intervalo $[t_1, t_2)$ se expresa por:

$$\frac{S(t_1) - S(t_2)}{(t_1 - t_2) S(t_1)} \quad (2.1)$$

De forma general, tomando el intervalo como $[t, t + \Delta t)$, en la expresión (2.1) toma la siguiente forma:

$$\lambda(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

Asumiendo Δt bien pequeño, $\lambda(t)$ representa la tasa de falla instantánea en el tiempo t condicional a la sobrevivencia a un tiempo t . Se observa que las tasas de falla son números positivos. (Colosimo & Ruiz Giolo, 2006) Menciona que: Una tasa de falla $h(t)$ es útil para describir una distribución de vida de algunos pacientes en aplicaciones médicas. Esta describe la forma en que la tasa instantánea de falla cambia con el tiempo.

Una función de riesgo para T está definida como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \cdot \frac{P(t \leq T < t + \Delta t \mid T > t)}{\Delta t} \quad (2.2)$$

Desarrollando la probabilidad condicional de la expresión anterior (2.2) se obtiene:

$$h(t) = \lim_{\Delta t \rightarrow 0} \cdot \frac{1}{\Delta t} \frac{P(t \leq T < t + \Delta t)}{P(T > t)} \quad (2.3)$$

Del cálculo de probabilidades y del hecho que $P(T \geq t) = S(t)$, se tiene que:

$$h(t) = \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \cdot \frac{F(t + \Delta t) - F(t)}{\Delta t} \quad (2.4)$$

De la definición de derivada se tiene que la expresión (2.4) se convierte en:

$$h(t) = \frac{f(t)}{S(t)} \quad (2.5)$$

Con esto $h(t)$ es una expresión que relaciona la función de sobrevivencia y la función de riesgo. Si se integra la expresión (2.5), se obtiene la Función de Riesgo acumulada:

$$H(t) = \int_0^t h(s) ds = \int_0^t \frac{f(s)}{S(s)} ds \quad (2.6)$$

Puede escribirse como $H(t) = \int_0^t \frac{f(s)}{1-F(s)} ds$, haciendo sustitución $u = 1 - F(s)$, y $du = -f(s)ds$, se tiene que la expresión (2.6) se convierte en $H(t) = \int \frac{-du}{u}$, de donde (Colosimo & Ruiz Giolo, 2006) Señala algunas relaciones matemáticas importantes entre las funciones referenciadas anteriormente, Luego:

$$h(t) = \frac{f(t)}{S(t)} = - \frac{d}{dt} (\ln S(t)) \quad (2.6a)$$

$$H(t) = \int_0^t h(u)du = - \ln S(t) \quad (2.6b)$$

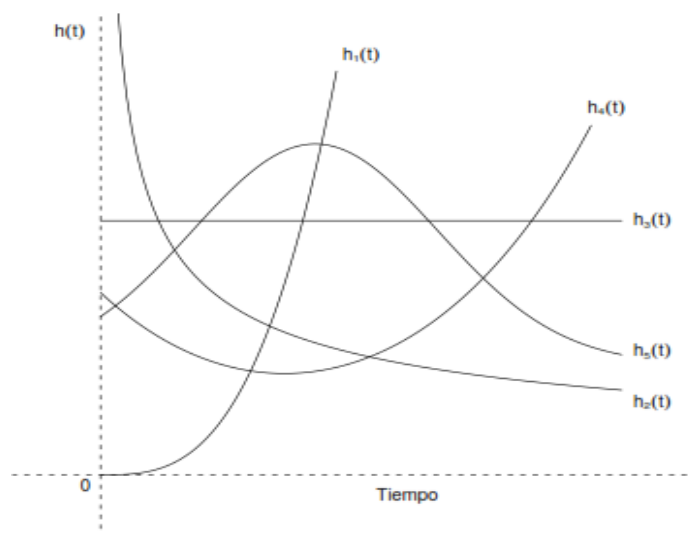
$$S(t) = \exp\{-H(t)\} = \exp\left\{- \int_0^t h(u)du\right\} \quad (2.6c)$$

Para ilustrar las funciones de riesgo mencionados por (Godoy, 2009) se presentan algunos escenarios en la gráfica 1(a), por ejemplo, pacientes con leucemia que no responden al tratamiento tienen una tasa de riesgo creciente $h_1(t)$, $h_2(t)$ es una función de riesgo decreciente que puede indicar el riesgo de soldados heridos por bala que fueron sometidos a cirugía. El peligro principal es la operación misma y este peligro decrece si la cirugía es exitosa. Una función de riesgo constante como en $h_3(t)$ es el riesgo de individuos saludables entre 18 y 40 años de edad cuyos riesgos principales de muerte son los accidentes. La llamada curva de tina de baño $h_4(t)$ describe el proceso de vida humana, durante el periodo inicial el riesgo es alto (alta mortalidad infantil), subsecuentemente el riesgo permanece aproximadamente constante hasta un cierto tiempo, después del cual crece debido a fallas por deterioro. Finalmente, pacientes con tuberculosis tienen riesgos que se incrementan inicialmente, luego decrecen después de tratamiento. Este incremento y luego decremento se muestra en la función de riesgo $h_5(t)$.

En la gráfica 1(b) (Aguilar, 2009) ejemplifica una función de riesgo acumulada estimada para tres grupos de insecticidas por separado donde se puede apreciar que el riesgo es mayor para las cucarachas tratadas con el insecticida C mientras transcurre el tiempo. Cada salto de estas funciones escalonadas corresponde al riesgo de falla diario de cada cucaracha tratada con el correspondiente insecticida,

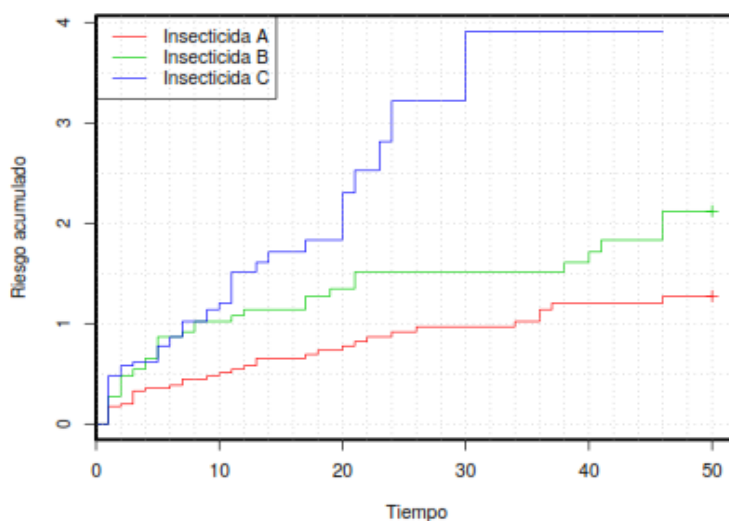
de manera que se puede interpretar con detalle la evolución del riesgo para cada insecticida por separado y en comparación con los demás mediante la longitud de los saltos de la función de riesgo acumulado.

Gráfica 1 (a) Representación de algunas funciones de Riesgo



(b)

Función de Riesgo Acumulado por insecticida



Cuando T es una variable aleatoria discreta, que toma valores t_j con $j = 1, 2, \dots$, la función de riesgo se define para los valores t_j y proporciona la probabilidad

condicional de falla de tiempo $t = t_j$, dado que el individuo estaba vivo antes de t_j , por lo que se tiene que:

$$\begin{aligned} h(t_j) &= P(T = t_j) / T \geq t_j \\ &= \frac{P(T = t_j)}{P(T \geq t_j)} \\ &= \frac{f(t_j)}{S(t_j-)} \end{aligned}$$

Donde $(t_j -)$ corresponde a un instante antes de t_j y por tanto:

$$P(T \geq t_j) = 1 - P(T < t_j) = S(t_j -) \neq S(t_j)$$

Y después de una serie de procedimientos matemáticos, según (Godoy, 2009), queda de la siguiente manera:

$$h(t) = \prod_{t_j < t} \frac{S(t_j)}{S(t_{j-1})}$$

2.2.2.5. Vida Media Residual

Otra función básica en el análisis de supervivencia según (Godoy, 2009) es la función de *vida media residual* al tiempo t denotada como $mrl(t)$ (por el nombre en inglés *mean residual life*). Para los sujetos de edad t , esta función mide la esperanza de tiempo de vida restante, o el tiempo esperado antes de la ocurrencia del evento de interés. Por tanto, queda definida por:

$$mrl(t) = E(T - t | T > t)$$

Para el caso continuo, por definición de esperanza condicional se tiene que:

$$\begin{aligned} E(T - t | T > t) &= \int_0^{\infty} (u - t) f(u | u > t) du \\ &= \int_0^{\infty} (u - t) \frac{f(u)}{S(t)} I_{(t, \infty)}(u) du = \int_t^{\infty} \frac{(u - t) f(u)}{S(t)} du \end{aligned}$$

Por lo cual la función de *Vida media residual* al tiempo t queda definida por:

$$mrl(t) = E(T - t/T > t) = \int_t^{\infty} \frac{(u - t)f(u)du}{S(t)}$$

Por lo que se puede apreciar que la vida media residual es el área bajo la curva de supervivencia a la derecha de t dividida entre $S(t)$. De tal modo que la vida media $\mu = E(T) = E(T - 0/T > 0) = mrl(0)$, es el área total de la curva de sobrevivencia, es decir,

$$\mu = E(T) = \int_0^{\infty} tf(t)dt = \int_0^{\infty} S(t)dt$$

De esta forma, para el caso de variables aleatorias continuas se tiene la relación

$$mrl(t) = \frac{\int_0^{\infty} (u - t)f(u)du}{S(t)} = \frac{\int_0^{\infty} S(u)du}{S(t)}$$

Para el caso discreto la función de Vida Media residual en el caso discreto como lo expresa (Godoy, 2009) está definida:

$$E(T - t/T > t) = \sum_{r=0}^{\infty} \frac{(t_r - r)P | T = t_r, T > t}{S(t)}$$

La cual queda expresada de la siguiente manera:

$$mrl(t) = \frac{(t_{i+1} - t) S(t_i) + \sum_{j>1} (t_{i+1} - t_j) S(t_j)}{S(t)} \text{ para } t_i \leq t < t_{i+1}.$$

Por otra parte (Borges, 2005) indica que para la selección de un modelo paramétrico, es conveniente usar la función de riesgo, pues de acuerdo a la información que el investigador tenga del fenómeno que causa la falla, puede determinar las características que el modelo debe seguir en la forma de la tasa de riesgo conforme avanza el tiempo. Por ejemplo, el modelo exponencial que presenta

riesgo constante, resultaría inadecuado para modelar el tiempo que tarda un individuo en morir cuando se le ha detectado una enfermedad terminal, pues en este caso, el riesgo debe ser claramente creciente. No obstante, puede haber situaciones donde se tenga evidencia para suponer que el riesgo puede ser constante en el tiempo, si fuera de interés modelar el tiempo que tarda en romperse la cuerda del violín de un concertista, puede ser que este dependa de la dificultad de las piezas que este tenga que tocar y el tiempo que invierta en practicar para perfeccionar el sonido, de modo que podría pensarse que la falla de la cuerda puede suceder en cualquier momento, independiente del tiempo que lleve colocada en el instrumento. Debido a los criterios mencionados para seleccionar los modelos paramétricos adecuados, se mencionan a continuación algunos modelos probabilísticos destacados: Exponencial, Weibull, Log- Normal, Gamma y Gamma Generalizada.

Estimación de la Función de Sobrevida

En el análisis de datos de sobrevivencia, debido a posibles censuras, las estadísticas que resumen la información pueden no tener las propiedades estadísticas deseadas, tales como insesgamiento. Por ejemplo, la media muestral ya no es un estimador insesgado de la media poblacional (del tiempo de sobrevivencia). De esta forma es necesario usar otros métodos para presentar los datos, ya que la función de sobrevivencia no puede ser obtenida directamente con argumentos probabilísticos debido a la censura o truncamiento en los datos, es importante el uso de algunos estimadores.

Autores como (Borges, 2005) mencionan distintas formas de estimar dicha función, entre los más conocidos son los basados en las tablas de vida, entre el que se incluye el estimador de Kaplan y Meier, siendo más práctico, ya que los tiempos de observación van contribuyendo a la estimación de la función de sobrevida, además otros estimadores son el Fleming – Harrington.

2.3.1. El Estimador de Kaplan Meier

El estimador de Kaplan-Meier (1958), de la función de sobrevivencia, es el estimador más común entre los paquetes estadísticos. El método de construcción del estimador es un método no paramétrico ya que no asume ninguna estructura para la función de distribución de probabilidad del tiempo de vida.

(Borges, 2005) Describe que el estimador de Kaplan Meier utiliza toda la información disponible, casos censurados y no censurados, para realizar la estimación de la función de sobrevivencia. El estimador en cualquier instante de tiempo es obtenido de la multiplicación de una secuencia de probabilidades condicionales de sobrevivencia estimadas. Cada probabilidad condicional estimada se obtiene del número de casos observados en riesgo y el número de “muertes” en un instante de tiempo y se calcula como $\frac{(n-d)}{n}$, donde n es el número de casos en riesgo en el tiempo (t_i) y d es el número de “muertes en (t_i) ”

El estimador de Kaplan Meier también es llamado límite producto, está definido por:

$$\widehat{S}_{KM} = \prod_{i:t_i < t} \left(1 - \frac{d_i}{n_i}\right) \quad (2.7)$$

Propiedades del estimador de Kaplan-Meier según (Colosimo & Ruiz Giolo, 2006):

- No necesita grandes muestras.
- Es consistente.
- Converge asintóticamente a un proceso de Gauss.
- Es el estimador de máxima verosimilitud de $\hat{S}(t)$.

Para construir intervalos de confianza y test de hipótesis para $S(t)$, es necesario validar con precisión el estimador de Kaplan – Meier. Este estimador así como los

otros, está sujeto a variaciones que deben ser descritas en forma de intervalos, una expresión para la varianza asintótica de este estimador está dada por la fórmula de Greenwood (1926):

$$Var(\widehat{S}_{KM}(t)) = [\widehat{S}_{KM}(t)]^2 \sum_{t_i < t} \frac{d_i}{n_j(n_j - d_j)}$$

El intervalo de confianza del 95% de escala plana (o de identidad), llamado así porque es obtenido de manera estándar al que se obtiene cualquiera de los intervalos de confianza, sin utilizar ninguna transformación, se obtiene mediante:

$$\widehat{S}_{KM}(t) \pm 1.96 \sqrt{\widehat{Var}(S_{KM}(t))}$$

Donde $\sqrt{\widehat{Var}(S_{KM}(t))} = \text{error estándar}$, de estimación del estimador de Kaplan-Meier.

2.3.2. Estimador de la Función de Riesgo Acumulada.

Este es llamado el estimador de Nelson-Aalen. Dado que la función de supervivencia se puede expresar en términos de la función de riesgo, $H(t) = -\ln S(t)$, si se desea estimar la función de riesgo acumulada, entonces se tiene que:

$$\widehat{H}(t) = -\ln \widehat{S}(t) \tag{2.8}$$

Donde $\widehat{S}(t)$ es el estimador de Kaplan-Meier de la función de supervivencia.

2.3.3. Comparación de curvas de sobrevida

Estadísticas comúnmente usadas pueden ser vistas como generalizaciones para datos censurados, de conocidos test-no paramétricos, entre las cuales se

encuentran: test de Logrank (Mantel, 1966) ya que es muy utilizado el análisis de sobrevida. Esta función proporciona métodos para comparar dos o más curvas de sobrevida, mediante tablas de contingencia donde algunas de las observaciones pueden ser censuradas y donde la agrupación global puede ser estratificada (Colosimo & Ruiz Giolo, 2006).

(Diaz, 2015) Menciona que La estadística Logrank puede extenderse para verificar la igualdad varias funciones de sobrevida $S_1(t), \dots, S_q(t)$. Los datos pueden ser dispuestos de en una tabla de contingencia de tamaño $2 \times q$ con d_{jg} unidades a las cuales ha ocurrido el evento y $n_{jg} - d_{jg}$ unidades expuestas al evento en la columna g . La tabla de datos es de la forma:

Tabla 1 Comparación de curvas de sobrevida

Evento	Grupos					Total
	1	...	g	...	q	
Ocorre	d_{j1}	...	d_{jg}	...	d_{jq}	d_j
No ocurre	$n_{j1} - d_{j1}$...	$n_{jg} - d_{jg}$...	$n_{jq} - d_{jq}$	$n_j - d_j$
Total	n_{j1}	...	n_{jg}	...	n_{jq}	n_j

La estadística $D_j = (d_{j1} - \mu_{d1}, \dots, d_{jq} - \mu_{djq})$ tiene media cero y matriz de varianzas y covarianzas $\Sigma_j = (\sigma_{j(gh)})$. Se conforma la estadística D sumando sobre todos los tiempos de falla:

$$D = \sum_{j=1}^k D_j,$$

Contiene la discrepancia entre los totales de eventos observados y esperados al momento t_j . Bajo el supuesto de que las k tablas de contingencia son independientes la matriz de varianzas-covarianzas del vector $\Sigma = \Sigma_1 + \dots + \Sigma_k$.

Una estadística para la verificación de la igualdad de las q funciones de sobrevida se basa en la estadística:

$$\mathcal{LR} = D^T \Sigma^{-1} D \quad (2.9)$$

Se tiene que para \mathcal{LR} en el caso de varios grupos para muestras de tamaño grande, bajo la hipótesis nula de igualdad de funciones de supervivencia, se distribuye como una chi-cuadrado con $q-1$ grados libertad.

Modelo de Riesgos Proporcionales

Existen diversos modelos de supervivencia que involucran covariables al incorporar la manera en que estas afectan el tiempo de falla del sujeto en estudio y uno de estos es el modelo de riesgos proporcionales. Si bien es cierto que el modelo de riesgos proporcionales es el más usado en bioestadística y en muchas disciplinas más, por ser el más entendido e implementado y los resultados que este proporciona al ser utilizado adecuadamente (García, 2014).

2.4. MODELO DE REGRESIÓN DE COX

El modelo de regresión descrito es conocido como modelo de regresión de Cox o modelo de riesgos proporcionales de Cox, permite estimar la relación que hay entre un conjunto de variables explicativas fijas X_1, X_2, \dots, X_n también conocidas como covariables, y la respuesta o tiempo de supervivencia; o más bien, con la función de riesgo $h(t; X)$, que es la tasa instantánea del suceso de interés.

En el modelo de regresión de Cox (1972): la función de tasa de riesgo del tiempo de falla de un modelo con vector de covariables dadas por X está definida de la siguiente manera:

$$h(t; X, \beta) = h_0(t) \exp(\beta^T X) \quad (2.10)$$

donde $h_0(t)$ es la función de riesgo base; y $\beta^T = (\beta_1, \beta_2, \dots, \beta_n)$ es el vector de parámetros de regresión.

El modelo de Cox se dice que es un modelo semi - paramétrico, debido a que incluye una parte paramétrica y otra parte no paramétrica.

- i) La parte paramétrica, $\exp(\beta^T X)$ llamada *función de riesgo relativo*, la cual está claramente especificada y describe los efectos relativos de los parámetros de regresión estimados sobre el riesgo.
- ii) La parte no paramétrica es $h_0(t)$, la función de riesgo base, que es una función arbitraria y no especificada.

Además, tal y como se ha mencionado anteriormente, el modelo de regresión de Cox también es conocido como modelo de riesgos proporcionales, ya que el cociente entre el riesgo para dos sujetos con el mismo vector de covariables es constante en el tiempo, es decir:

$$\frac{h(t; X_i)}{h(t; X_j)} = \frac{h_0(t) \exp(\beta^T X_i)}{h_0(t) \exp(\beta^T X_j)} = \frac{\exp(\beta^T X_i)}{\exp(\beta^T X_j)} = \exp(\beta^T (X_i - X_j)) \quad (2.11)$$

El cociente expresado en la ecuación (2.11) es conocido como *razón de riesgos relativos*.

2.4.1. Estimación de parámetros

El procedimiento más utilizado para realizar la estimación de los parámetros de un modelo de regresión es el procedimiento relacionado con la verosimilitud que básicamente consiste en estimar los parámetros que maximice el logaritmo natural de la verosimilitud, en el caso del modelo de Cox este procedimiento no funciona básicamente debido al desconocimiento de la función de riesgo base que aparece en el modelo.

Cox en 1972 desarrolla un método que permite estimar los parámetros sin necesidad de asumir una forma particular para la función de riesgo base, propone entonces usar una expresión llamada “*función de verosimilitud parcial*” que depende solamente de los parámetros de interés. Equivalentemente se puede considerar la función de riesgo acumulado base $H_0(t) = \int_0^t h_0(u)du$ o la función de sobrevivencia base $S_0(t) = \exp[-H_0(t)]$, en lugar de la función riesgo base. La función parcial se puede escribir como:

$$L(\beta) = \prod_{i=1}^n [h_0(t_i) \exp(\beta^T X_i)]^{\delta_i} \exp[-H_0(t_i) \exp(\beta^T X_i)] \quad (2.12)$$

Siendo δ_i el indicador de censura, $\delta_i = 0$ si la observación es censurada o, por el contrario $\delta_i = 1$. Utilizando la expresión (2.6c) la función de verosimilitud se puede expresar como:

$$L(\beta) = \prod_{i=1}^n [h_0(t_i) \exp(\beta^T X_i)]^{\delta_i} S_0(t_i)^{(\beta^T X_i)} \quad (2.13)$$

En el modelo de Cox no se supone una forma específica de $h_0(t)$ por lo que, no es posible emplear directamente el método (2.13) para obtener un estimador del vector de parámetros β . Pese a esto, Cox propuso la siguiente función de verosimilitud para estimar β :

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp \beta^T X_i}{\sum_{j \in R(t_i)} \exp \beta^T X_j} \right]^{\delta_i} \quad (2.14)$$

donde $R(t_i) = \{j : t_j > t_i\}$ es el conjunto que contiene los sujetos en riesgo en el tiempo t_i , es decir, que están en observación y aún no han presentado el evento.

Puesto, que δ_i es el indicador de censura y será 0 cuando la observación esté censurada, la función de verosimilitud parcial quedará:

$$L(\beta) = \prod_{i=1}^m \frac{\exp \beta^T X_i}{\sum_{j \in R(t_i)} \exp \beta^T X_j} \quad (2.15)$$

siendo i ahora los índices de los m tiempos de eventos observados, t_1, t_2, \dots, t_m .

La correspondiente función de log-verosimilitud parcial es:

$$\ln(L(\beta)) = \sum_{i=1}^m \left\{ \beta^T X_i - \ln \left[\sum_{j \in R(t_i)} \exp(\beta^T X_j) \right] \right\} \quad (2.16)$$

De modo que, la estimación de los parámetros β se obtiene maximizando la función de verosimilitud parcial (2.15) o, de forma equivalente, maximizando la función de log-verosimilitud parcial (ecuación (2.16)) para $\beta_1, \beta_2, \dots, \beta_n$, sin necesidad de estimar la función de riesgo base $h_0(t)$.

Sin embargo, cuando los datos contienen tiempos observados empatados, la verosimilitud parcial (2.15) puede llevar un tiempo de computación considerable. Por esta razón, cuando se tienen datos con empates, se utilizan aproximaciones para la función de verosimilitud parcial. Una aproximación de esta fue sugerida por Breslow(1974). Esta aproximación considera que los d_i eventos al tiempo t_i son distintos y ocurren secuencialmente que vienen dados por la siguiente fórmula:

$$L(\beta) = \prod_{i=1}^m \frac{\exp \beta^T X_i}{\left[\sum_{j \in R(t_i)} \exp \beta^T X_j \right]^{\delta_i}} \quad (2.17)$$

Y el logaritmo de la Verosimilitud

$$\ln(L(\beta)) = \sum_{i=1}^m \left\{ \beta^T S_i - \delta_i \ln \left[\sum_{j \in R(t_i)} \exp(\beta^T X_j) \right] \right\} \quad (2.18)$$

donde $S_i = j \in R(t_i) X_j D t_i = \{i_1, \dots, i_d\}$ es el conjunto de etiquetas de los individuos que experimentan el evento a tiempo t_1 y d_i es el número de eventos a tiempo t_i .

Función de supervivencia para el modelo riesgos proporcionales.

Una vez que ya se tienen los estimadores de β , $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n)$ se obtiene el estimador de h_j la función de riesgo inicial entre t_{j-1} y $t_j (t_0 := 0)$ con $j = 1, 2, \dots, d$, se muestra a continuación:

$$\hat{h}_j = \frac{1}{t_j - t_{j-1}} \frac{1}{\sum_{k \in R(t_j)} \exp(\beta^T X_k)} \quad (2.19)$$

2.4.2. Contraste de hipótesis del modelo de riesgos proporcionales

Una vez que se ha ajustado un modelo de Cox, se verifica que sean significativas las variables del modelo. Para ello, existen tres pruebas que se encargan de validar la hipótesis de que los parámetros del modelo de Cox son significativos. En ellas, se considera el vector de parámetros estimados $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_1, \dots, \hat{\beta}_n)^T$ y la matriz de información evaluada en β , definida como:

$$I(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial \beta_k^2}, k = 1, \dots, n.$$

La primera de ellas es la prueba de Wald, prueba de razón de verosimilitudes, prueba de puntajes, y estas son asintóticamente equivalentes.

1. **Test de Wald:** Este test es, quizás el más natural debido a que proporciona un contraste por variables en vez de una medida de significación global. El estadístico de contraste se define mediante:

$$X_W^2 = (\hat{\beta} - \beta_0)^T I^{-1}(\hat{\beta}) (\hat{\beta} - \beta_0) \quad (2.10)$$

Donde $I(\beta)$ es la matriz de varianzas y covarianzas estimada. Esta prueba tiene como hipótesis nula $H_0 : \beta = \beta_0$ y sigue la distribución Chi cuadrado con n grados de libertad si H_0 es cierta para muestras grandes.

2. **Test de Razón de Verosimilitud:** es el que presenta mayor confiabilidad, este se define como:

$$2\{\ln(L(\beta_0)) - \ln(L(\hat{\beta}))\} \quad (2.21)$$

Donde β_0 son los valores iniciales de los coeficientes y $\hat{\beta}$ es la solución luego de ajustar el modelo.

3. **Test de Puntajes (score test):** El tercer contraste es el conocido como test de los puntajes, definido como $U(\beta) = (U_1(\beta), U_2(\beta), \dots, U_n(\beta))^T$ donde U es el vector de derivadas del $\ln(L(\beta))$. Para muestras grandes, cuando H_0 es cierta, $U(\beta)$ tiene distribución asintótica normal con vector cero por media y matriz de covarianzas dada por $I(\beta)$. Teniendo como hipótesis nula $H_0 : \beta = \beta_0$, la estadística de prueba está dada por:

$$X_{sc}^2 = U(\beta_0)^T I^{-1}(\beta_0) U(\beta_0) . \quad (2.22)$$

La cual tiene una distribución Chi-cuadrado con n grados de libertad.

Y al tener en cuenta el criterio con los valores “pequeños” en el p-valor, se tiene evidencia de que, bajo las pruebas mencionadas anteriormente, los coeficientes del

modelo son significativamente distintos de cero, y por tanto, se podría considerar que el modelo tiene sentido para las variables explicativas consideradas.

Interpretación del modelo

Cuando el modelo de riesgos proporcionales es utilizado, los coeficientes de las variables explicativas en el modelo, pueden ser interpretados como el logaritmo del cociente de la función de riesgo y la función de riesgo inicial. La interpretación de los parámetros correspondientes a los diferentes tipos de variables explicativas del modelo de riesgos proporcionales depende si éstas son variables o factores.

Si el coeficiente de regresión (β) corresponde a una variable, la cantidad $\exp(\beta)$ es el cambio en la función de riesgo por cada unidad que se incremente la variable, dado que las otras covariables están fijas, esto se puede escribir como sigue:

$$\begin{aligned} & \frac{h(t, x_1, \dots, x_i + 1, \dots, \beta_1, \dots, \beta_p)}{h(t, x_1, \dots, x_i, \dots, \beta_1, \dots, \beta_p)} \\ &= \frac{h_0(t) \exp \{ \beta_1 x_1 + \dots + \beta_i (x_i + 1) + \dots + \beta_p x_p \}}{h_0(t) \exp \{ \beta_1 x_1 + \dots + \beta_i x_i + \dots + \beta_p x_p \}} \\ &= \exp \{ \beta_i \}. \end{aligned}$$

Si el coeficiente de regresión β corresponde a un factor, la interpretación es de la siguiente manera: suponga que se tiene un factor A con a niveles, y sean β_2, \dots, β_a los coeficientes de regresión de los correspondientes niveles, considerando que el primer nivel del factor A es cero. La cantidad $\exp\{\beta_i - \beta_j\}$ es el cociente de la función de riesgo para el sujeto con nivel i y nivel j de las variables explicativas ($i, j = 2, \dots, a$), dado que las otras variables explicativas toman valores iguales. La

cantidad $\exp\{\beta_i\}$ corresponde al cambio relativo en la función de riesgo para los sujetos con las variables explicativas con nivel $i = (i=2, \dots, a)$ y nivel 1.

Para efectos posteriores en la interpretación de los parámetros (Osorio & Castillo, 2011), mencionan que el signo positivo o negativo de los coeficientes (β) representa el incremento (o disminución) en la probabilidad de desertar/graduarse cuando se incrementa en una unidad la variable explicativa (si ésta es continua), manteniendo todas las demás constantes. En el caso de las variables dicotómicas, el parámetro(β) representa la variación en el riesgo de deserción o propensión a graduarse cuando un individuo se encuentra en una categoría con respecto a la categoría base.

2.4.3. Método de selección del Modelo

(García, 2014) menciona que el principal objetivo a la hora de establecer un modelo es buscar aquel que contenga solamente los efectos principales, es decir, con las covariables que han resultado significativas de la regresión de Cox. Para determinar cuáles de estas variables aportan más en el modelo y que, a su vez, no estén relacionadas, existen diferentes métodos de selección de variables, los cuales son:

i. Forward (hacia adelante): en este método se inicia el proceso con un modelo vacío, solo con el término independiente. Se ajusta un modelo con el método de máxima verosimilitud y se calcula el estadístico Chi-cuadrado con el p-valor de incluir cada variable por separado. Luego, se selecciona el modelo con la variable más significativa, es decir, que tiene un p-valor $p < 0,05$. De nuevo, se ajusta el modelo con las variables seleccionadas y se calcula el p-valor resultante de añadir cada una de las variables que no han sido seleccionadas por separado. Tras escoger el modelo con la variable más significativa, se repiten estos pasos hasta que no queden variables significativas por incluir.

ii. Backward (hacia atrás): consiste en empezar con un modelo que contiene todas las variables candidatas e ir eliminando, una a una, cada covariable, a la vez que se calcula la pérdida de ajuste al eliminar. Se omite del modelo la variable menos significativa, esto es, aquella que mayor p-value tenga y se repiten estos pasos hasta que todas las variables incluidas sean significativas y no pueda eliminarse ninguna sin que se pierda ajuste.

iii. Stepwise: este método es una combinación de los procesos forward y backward. Se puede comenzar, o bien, con el modelo vacío, o bien, con el modelo completo, pero en cada paso se exploran las variables incluidas por si deben salir y las no seleccionadas por si deben entrar en el modelo. Se repiten estos pasos hasta que todas las variables incluidas sean significativas y no entre ni salga ninguna más.

En este trabajo, se utiliza el método forward (hacia adelante) y el criterio de AIC(Akaike's Information Criterion) que es mencionado en seguida:

Criterio AIC

(Akaike's Information Criterion). El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo. AIC maneja un trade-off entre la bondad de ajuste del modelo y la complejidad del modelo. Se da una estimación relativa de la información pérdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. En general AIC es:

$$AIC = 2k - 2 \ln(L) \quad (2.23)$$

donde k es el número de parámetros en el modelo estadístico , y L es el máximo valor de la función de verosimilitud para el modelo estimado.

Dado un conjunto de modelos candidatos para los datos, sirve para seleccionar el mejor modelo dentro de un conjunto de estos obtenidos con los mismos datos. Por lo tanto AIC no sólo recompensa la bondad de ajuste, sino también incluye una penalidad, que es una función creciente del número de parámetros estimados. Esta penalización desalienta el sobreajuste (aumentando el número de parámetros libres en el modelo mejora la bondad del ajuste, sin importar el número de parámetros libres en el proceso de generación de datos).

2.4.4. Evaluación del modelo

Una de las ventajas del enfoque del análisis de sobrevivencia es la posibilidad de efectuar la bondad de ajuste del modelo de riesgos proporcionales mediante un análisis de residuos. De modo que un residuo es el valor calculado, para cada individuo, como la diferencia entre el valor de sobrevivencia observado y el valor estimado por la ecuación de regresión. Cuanto mayor es esa diferencia mayor será el valor del residuo, con su signo correspondiente.

El análisis de residuos en cualquier modelo estadístico según (García, 2014). Sirve para verificar la adecuación del modelo ajustado por medio de inspección de gráficos. De manera que, los residuos en el modelo de Cox pueden ser utilizados para:

- Descubrir la forma funcional correcta de un predictor continuo.
- Identificar los sujetos que están pobremente predichos por el modelo.
- Verificar el supuesto del modelo de regresión de Cox.

Dentro de este hay seis tipos de residuos: residuos de Cox-Snell, de martingala, de desvíos (deviance), de puntajes (score) y de Schoenfeld. De estos residuos pueden derivarse otros, como los df_{betas} , de los cuales se muestran los de interés para abordar este estudio:

I. Residuos de Cox-Snell: Este tipo de residuos, desarrollados por Cox & Snell, sirven para evaluar el ajuste global del modelo planteado. Si el modelo de regresión de Cox definido por la ecuación [10] es adecuado, entonces las estimaciones del tiempo de supervivencia del modelo planteado vienen dadas por un estimador de la función de supervivencia $\widehat{S}_i(t)$, el cual debe ser muy similar al valor verdadero de $S_i(t)$. Para evaluar esto, se calculan los residuos para los n individuos en estudio, de modo que:

$$r_{cs_i} = \widehat{H}_0(t_i) \exp(\widehat{\beta}^T X_i), i = 1, \dots, n.$$

en donde $\widehat{H}_0(t_i)$ es el estimador de la función de riesgo acumulado base de Breslow definido por:

$$\widehat{H}_0(s) = \int_0^s \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) \exp(\widehat{\beta}^T X_i(s))}.$$

El cual se basa en un proceso de conteo $N_i \equiv \{N_i(t), t \geq 0\}$, que para el i -ésimo sujeto es el número de eventos observados hasta el tiempo t y donde $Y_i(s)$ es un proceso 0 – 1 que indica si el i -ésimo sujeto está en riesgo en el tiempo t .

Para probar si los residuos de Cox-Snell están o no aproximadamente distribuidos de forma exponencial, se construye su gráfico de residuos y si estos residuos están distribuidos de forma exponencial, entonces una estimación de la tasa de riesgo basada en r_{cs_i} representada frente a los residuos r_{cs_i} debería tender a una línea recta que pasa por el origen con pendiente igual a la unidad. Es decir, el riesgo acumulado $H_r(r_{cs_i})$ frente a los residuos r_{cs_i} , debería ser aproximadamente una línea recta que pasa por el origen con pendiente igual a 1.

II. Residuos de Schoenfeld: estos residuos se definen como la matriz:

$$S_{ij}(\beta) = X_{ij}(t_i) - \bar{X}_j(\beta_i, t_i)$$

la cual contiene una fila por evento ocurrido y una columna por covariable, donde i y t_i son los individuos y el tiempo de ocurrencia del evento, respectivamente.

Los residuos de Schoenfeld son útiles para la verificación del supuesto de riesgos proporcionales en el modelo de Cox. Estos residuos, representados frente al tiempo observado de supervivencia deben repartirse aleatoriamente alrededor de cero, siempre que el modelo de Cox sea correcto. El estudio también puede ser abordado numéricamente. Como alternativa a la hipótesis de riesgos proporcionales, Therneau y Gramsch (2000) consideran que los coeficientes de regresión están dados por funciones de tiempo de la forma $\beta(t) = \beta + \theta g(t)$, siendo $g(t)$ una función suave. En donde se contrasta la hipótesis nula : $H_0: \theta = 0$, el modelo de riesgos proporcionales es correcto. Los p-valores en la tercera columna indican que la hipótesis nula puede ser aceptada con bastante contundencia en todos los casos.

	θ	Chisq	P
β			

III. Residuos dfbeta: sirven para determinar la influencia de cada observación en la estimación de los coeficientes de regresión. Este residuo calcula el cambio aproximado en el j -ésimo coeficiente (es decir, la j -ésima covariable) si la observación i -ésima se elimina del conjunto de datos y se vuelve a estimar el modelo sin esta observación. Así para el individuo i el valor dfbeta correspondiente a la variable j es el siguiente:

$$dfbeta_i = \beta_i - \beta_j(\text{excluyendo } i).$$

Los valores dfbeta pueden estandarizarse dividiendo por el error estándar del coeficiente correspondiente. En su representación gráfica se suelen mostrar los valores de los residuos dfbeta estandarizados para cada covariable del modelo frente a los índices de individuo (número de orden). Si la supresión de una observación hace que el coeficiente incremente, el residuo dfbeta es negativo y viceversa.

2.4.5. Pronóstico en el modelo de Riesgos Proporcionales

Teniendo en cuenta que la estimación o ajuste del modelo de riesgos proporcionales está dado por la expresión $h_i(t, X, \beta) = h_0(t) \exp(X\beta^T)$. Las dos componentes del modelo $h_0(t)$ (función de riesgo inicial) y $\exp(X\beta^T)$ pueden ser estimadas por separado, se estiman primero los componentes del vector de parámetros y a partir de ellos se construye el estimador de la función de riesgo inicial. Nótese que la función de riesgo inicial depende del tiempo mientras que la segunda componente depende únicamente de las covariables. Sin embargo, en el modelo de riesgos proporcionales de Cox se debe tener en cuenta que los riesgos para dos conjuntos diferentes de valores de los covariables conservan la misma proporción a lo largo del tiempo, de ahí su denominación.

El exponente del modelo de Cox se denomina *índice pronóstico* (P_i), Un índice pronóstico centrado igual a cero corresponde a un estudiante “base”. La diferencia entre los índices pronósticos de dos estudiantes permite estimar su riesgo relativo (RR), como se muestra a continuación:

$$RR = \frac{h_i(t, X_i, \beta)}{h_j(t, X_j, \beta)} = \frac{h_0(t) \exp(P_{ICindividoi})}{h_0(t) \exp(P_{ICindividoj})} = \exp(P_{ICindividoi} - P_{ICindividoj})$$

3. DISEÑO METODOLÓGICO

Esta sección tiene como propósito presentar el diseño metodológico para desarrollar el proyecto, en primer lugar se presenta diseño muestral, luego se describen las variables objeto de estudio, la fuente de información que permite establecer de una manera clara y ordenada los estudiantes matriculados de los diferentes programas de Administración de la facultad seccional Duitama, y finalmente las herramientas para el análisis de datos.

3.1. DISEÑO MUESTRAL

Población

Constituida por el total de estudiantes matriculados en primer semestre de Administración Industrial, Turística y Hotelera, Empresas Agropecuarias que ingresaron en 2004 primer semestre hasta el segundo semestre de 2009 .

De acuerdo a la información suministrada por la Oficina de Control y Registro Académico – SIRA, esta población tiene un total de 1489 estudiantes distribuidos de la Siguiente manera:

- Administración Industrial:536
- Administración Turística y Hotelera:489
- Administración de Empresas Agropecuarias:464

Con el fin de reducir los costos, una mayor rapidez y exactitud se seleccionó una muestra representativa aleatoria sin reemplazo. Al aproximar las características poblacionales mediante estimadores basados en la muestra se comete un error, error que mide la representatividad de dicha muestra (Pérez, 2000), razón por la cual se prefijará un error de muestreo mínimo para determinar el tamaño de

muestra. A continuación se precisa el tipo de muestreo, marco muestral, unidades de muestreo, error de muestreo y nivel de confianza.

Tipo de Muestreo

Se utilizó para este caso es un Muestreo Aleatorio Simple sin Reposición M.A.S. Este consiste en la selección de un subconjunto de elementos de la población en forma aleatoria sin reposición, los cuales tienen la misma probabilidad de ser seleccionados y el orden no interviene. Debido a que el procedimiento de selección es con probabilidades iguales, todas las muestras son equiprobables, es decir, el M.A.S es un método de selección de muestras en las cuales las unidades se eligen individual y directamente por medio de un proceso aleatorio en el que cada unidad no seleccionada tiene la misma oportunidad de ser elegida que todas las otras unidades en cada extracción de la muestra. (Lininger & Warwick, 1978)

Marco muestral

De la base de datos suministrada por el SIRA, se toma el listado de estudiantes matriculados en primer semestre pertenecientes a los programas de Administración de la UPTC seccional Duitama, de la cual de forma aleatoria se selecciona la muestra listándolos mediante el código estudiantil y, en algunos casos con el número de identificación, (cedula de ciudadanía), de cada estudiante seleccionado en la muestra. Posteriormente se cotejan y complementan las variables con el SPADIES, para finalmente formar la base de datos a trabajar por programa académico.

Unidad de muestreo

Para el caso que nos compete, la unidad elemental de muestreo corresponde al estudiante que pertenece a cada programa de administración de la UPTC – Duitama durante el periodo de tiempo especificado anteriormente.

Selección del tamaño de la muestra

Los parámetros para calcular el tamaño de la muestra son:

Tabla 2 Parámetros para determinar tamaño de muestra

Parámetro	Descripción	Programa Académico		
		A.I	A.E.A	A.T.H
N	Total de estudiantes matriculados, de cada programa de administración de la UPTC Duitama	536	464	489
n	El tamaño de la muestra calculado	178	155	173
e	Error máximo aceptable, e= 0.06 ó 6%			
P	Es la proporción que esperamos encontrar= 0.5			
1 - p	0.5			
Z	Valor correspondiente de la curva normal con una confiabilidad del 95% = 1,96			

Fuente: Autor

Para este caso de la estimación se usará una proporción, se estima el tamaño de la muestra según el supuesto finita de la población, obteniendo:

$$n = \frac{N \cdot Z^2 \cdot (p) \cdot (1 - p)}{(N - 1) \cdot e^2 + Z^2 \cdot (p) \cdot (1 - p)}$$

Esta se utilizará una por cada programa para obtener una muestra representativa de la siguiente manera para cada programa académico:

Administración Industrial (A.I)

$$n = \frac{(536) \cdot (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}{(535) \cdot (0,06)^2 + (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}$$

$$n = \frac{514,7744}{2,8756}$$

$$n = 178,3447 \approx 178$$

Con lo cual el tamaño de muestra es de 178 estudiantes de AI.

Administración de Empresas Agropecuarias (AEA)

$$n = \frac{(464) \cdot (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}{(463) \cdot (0,06)^2 + (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}$$

$$n = \frac{445,6256}{2,876}$$

$$n = 154.9464 \approx 155$$

Así que son seleccionados son 155 estudiantes de AEA.

Administración Turística y Hotelera (ATH)

$$n = \frac{(489) \cdot (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}{(488) \cdot (0,06)^2 + (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}$$

$$n = \frac{469.6356}{2,7172}$$

$$n = 172,8380 \approx 173$$

De modo que la muestra es de 173 estudiantes de ATH.

(En los anexos: A, B, C está el listado de los estudiantes seleccionados por programa y los detalles de las Sentencias en “**R**” para la selección de la muestra)

3.2. RECOLECCIÓN DE LA INFORMACIÓN

Está se realizó apoyándose en los recursos institucionales que permitieron complementar la información que se requería para abordar este estudio, la primera fuente de información empleada es el SIRA, posteriormente se utilizó SPADIES, a continuación una breve descripción de estas:

Sistema de Información de Registro Académico (SIRA): es el sistema académico que maneja la UPTC y contiene los siguientes módulos: aspirantes, recibos, exenciones y deudas, expedición de certificaciones, control de calificaciones, inscripción de asignaturas, horarios, graduaciones, docentes y evaluación a docentes, planes de estudio, datos estadísticos, auditoria, módulos en internet, de los cuales se obtuvieron obtener algunas de las variables consideradas en el estudio, como por ejemplo: Número de materias perdidas, Número de materias

repetidas, Fechas de inicio del programa, Graduación, Estado del estudiante, entre otras.

SPADIES: Es una herramienta desarrollada a través de un convenio con el MEN, y que busca encontrar causas para deserción académica de la Universidad. Esto lo hacen todas la Universidades del País, permitiendo al estado acercarse a la realidad de la deserción en Colombia, por esto y conociendo que de manera general la información de los estudiantes que ingresan a la UPTC actualmente, se encuentran sincronizados los datos desde el primer semestre de 1998, Está disponible en el link <http://aplica.uptc.edu.co/News/Paginas/SPADIESArchivosdeConsulta.aspx>, con esto facilitó el acceso a algunas variables que los estudiantes registran en el momento de presentar el examen de estado en el grado once “ICFES” o actualmente conocido como las pruebas “Saber 11” como son: Nivel educativo de la Madre, Número de Hermanos, Poseer Vivienda, entre otras que no tiene el SIRA. Para este caso se consultó esta base de datos seleccionado a la seccional Duitama que corresponde al código: IES 1107 donde la información de SPADIES corresponde a la versión. 2.8.2 (Actualizada a Agosto de 2015 con la información de matrícula del semestre 1-2015).

3.3. VARIABLES OBJETO DE ESTUDIO

Variables es una característica que tiene en común todos los elementos de un conjunto de personas o cosas de tal manera que, al medirla en los elementos de ese conjunto, se obtienen valores diferentes e impredecibles (Lerma, 2009). Las variables que simplemente clasifican o caracterizan sectores de la población se conocen como *atributo*, *variable categorica* o *cualitativa* (Ospina, 2001). Se dice que es una variable categorica *nominal* cuando está no establece ninguna relación de orden entre sus categorias, cuando existe esta relación de orden entre categorias se llaman variables categoricas *ordinales*.

Cuando las características pueden ser medidas a partir de valores que son números pero que además estos conservan sentido práctico de sus propiedades intrínsecas (cuantificar la diferencia entre dos posibles valores, sumarlos, etc.), estas características se conocen como *variables cuantitativas*.

Cuando una variable numérica se caracteriza por la propiedad de que para dos posibles valores de ella solamente hay un número finito de posibles valores intermedios se llama *variable cuantitativa discreta*. En caso que entre dos posibles valores de ella cualquier valor intermedio es también un valor posible de la variable se llama *variable cuantitativa continua*.

A continuación se hará la descripción de las variables que son objeto de medición en este proyecto, indicando el tipo de variable y su escala de medición.

Tabla 3 Variables objeto de estudio

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Tipo de Vivienda	Tipo de Vivienda (Posee o Carece) en el momento de presentar el ICFES	Variable cualitativa
Género	Masculino= H , Femenino= F	Variable cualitativa
Ingresos Familiares	Clasificados en tres categorías, según el periodo de ingreso del estudiante: Alto Medio Bajo	Variable cualitativa
No. De hermanos	Nº de Hermanos en el momento de presentar el ICFES.	Variable cuantitativa discreta
Puntaje estandarizado ICFES	En escala de 1 a 100	Variable cuantitativa discreta
Edad de Ingreso al Programa Académico	Edad en años cumplidos del Estudiante en el momento de ingresar a primer semestre en la UPTC.	Variable cuantitativa continua
Número de Materias perdidas	Asignaturas pérdidas durante los semestres cursados en la Universidad.	Variable cuantitativa discreta
Número de materias repetidas	Asignaturas repetidas durante los semestres cursados en la Universidad.	Variable cuantitativa discreta
Desertó	Abandono (Si o No) del programa de estudio. Este abandono puede ser por no renovar matrícula por tres semestres académicos consecutivos Artículo 42 o por causas académicas contempladas en el Artículo 80 literales A, B, C, D Y E del reglamento estudiantil A130/98.	Variable cualitativa
Permanencia en Semestres	Número total de semestres desde el inicio del programa académico hasta desertar o graduarse	Variable cuantitativa discreta

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Estado del Estudiante	Matriculado Desertor Graduado Terminación Académica Matriculado -Terminación Académica No matriculado	Variable cualitativa
Nivel Educativo de la Madre	Se refiere al máximo nivel de estudios alcanzados por la Madre a la hora de presentar el examen ICFES, estos son: Básica Primaria Básica Secundaria Técnico Universitario	Variable cualitativa
Trabajaba al presentar el examen de estado	Define si el estudiante laboraba cuando presentó el ICFES, toma los valores: Si o No	Variable cualitativa
Tiempo hasta la Deserción	Medido en Semestres Académicos (16 semanas)	Variable cuantitativa discreta
Tiempo Hasta la Graduación	Medido en Semestres Académicos	Variable cuantitativa discreta

Fuente: Autor

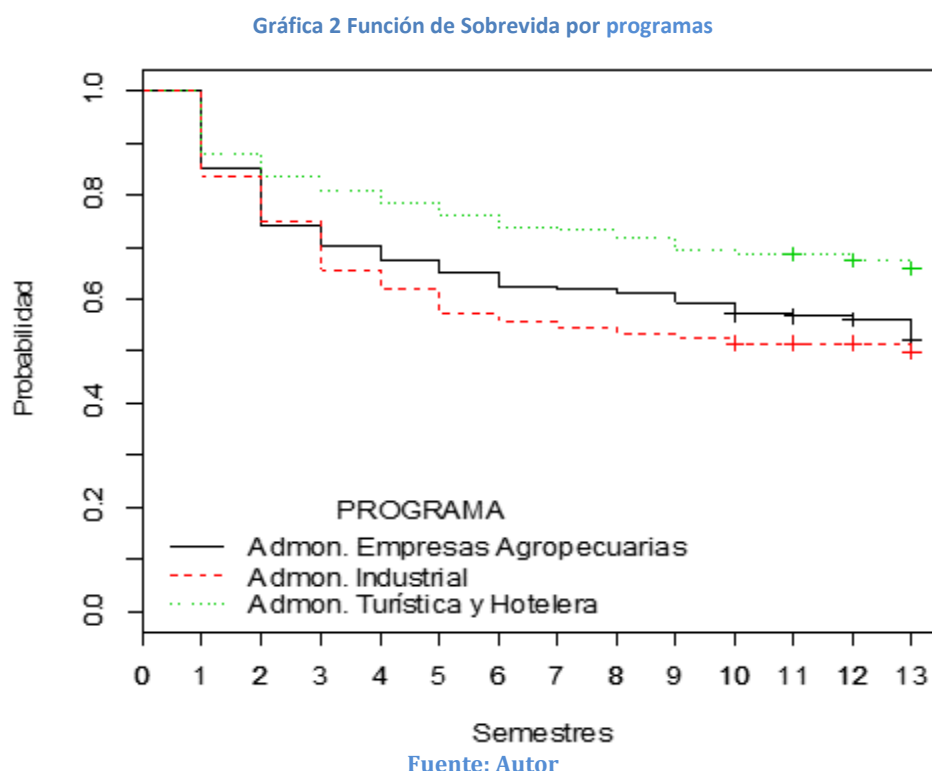
3.4. HERRAMIENTAS PARA EL ANÁLISIS DE DATOS

En el desarrollo del análisis de sobrevivencia llevado a cabo en este trabajo, se hace uso del lenguaje **R**, por la gran cantidad de funciones y métodos estadísticos implementados en él y, por el hecho, de ser utilizado por la gran mayoría de la comunidad científica. **R** es un lenguaje de programación de alto nivel diseñado de licencia libre, principalmente, para el estudio de grandes datos y estadísticos, puesto que abarca una amplia gama de técnicas estadísticas que van desde los modelos lineales a las más modernas técnicas de clasificación, pasando por los test clásicos y el análisis de series temporales. Por esto se utilizó **R** versión 3.2.2 (2015-08-14).

Entre las bibliotecas en **R** que permiten llevar a cabo un análisis de supervivencia, pero la más utilizada y la empleada en este trabajo, es “survival”. Dicha biblioteca es seleccionada por ser capaz de soportar datos que presentan censura. Además, contiene numerosas rutinas utilizadas para el desarrollo de este estudio.

Además en este software se realizó la selección de la muestra, análisis bivariado (ver Anexo E) de las variables objeto de estudio, además en el anexo E se presenta la sintaxis para cada programa en donde se efectúa el análisis de sobrevivencia.

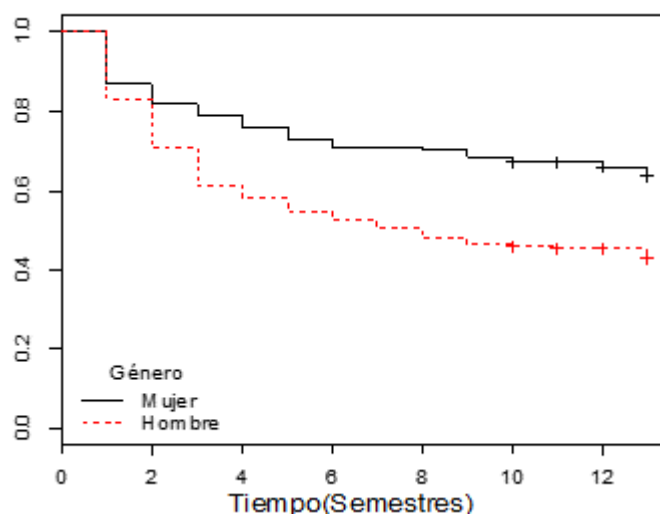
Antes de comenzar el análisis por cada Programa se decidió efectuar un análisis conjunto de las funciones de sobrevivencia. Lo anterior con el fin de juzgar la hipótesis de que si la función de sobrevivencia era la misma para los tres programas de administración, se encontró que bajo la prueba de Long-rank (ecuación (2.9)), ya que el p-valor obtenido es 0.00438, y es mucho menor que el nivel de significancia de 5%, se rechaza la hipótesis nula que indica igualdad de curvas de supervivencia, y por tanto se puede afirmar que la sobrevivencia por cada programa académico no tiene el mismo comportamiento, como se evidencia en la gráfica 2 :



Existen diferencias significativas en el comportamiento de la deserción entre AEA, AI y ATH, a partir del primer semestre los estudiantes de ATH tienen mayor probabilidad de no desertar, y tienen menor probabilidad de no desertar los

estudiantes de AI, apareciendo en medio AEA. Por lo cual se construirá un modelo por cada programa académico.

Gráfica 3 Función de sobrevivencia por género en las Administraciones



Fuente: Autor

Los porcentajes de censura para cada programa se muestran en la tabla 4:

Tabla 4 Resumen de censura por programas

Programa	N° Total	N° de Eventos	Censurado	
			N	Porcentaje
Admón. Empresas Agropecuarias	155	72	83	53,5 %
Admón. Industrial.	178	88	90	50,6 %
Admón. Turística y Hotelera.	173	58	115	66.5 %
Global	506	218	288	57 %

Fuente: Autor

Donde, *N° total*: se refiere a la cantidad de estudiantes seleccionados en la muestra e que pertenecen al programa que se indica, además donde el N° de eventos señala la cantidad de estudiantes desertores durante el periodo de análisis corresponde a 218 estudiantes; los casos censurados hacen referencia a estudiantes que no presentan el evento.

4. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS

En este capítulo se describe el proceso que se empleó para elaborar este modelo de sobrevivencia, que incluye un análisis de datos (de forma univariada y bivariada), posteriormente se referencian los supuestos del modelo y se muestra el modelo para la deserción y en seguida el de graduación, en donde para cada uno, se emplean el estimador de la función de sobrevivencia Kaplan-Meier, función de riesgo acumulada, adicionado a esto se seleccionan las variables, luego se estima el modelo de regresión de Cox, y por último se validan los supuestos de este.

4.1. DESCRIPCIÓN DE LOS DATOS

Esta sección muestra, en primer lugar, una descripción de forma univariada para cada una de las variables de interés. Algunas de ellas se refieren a aspectos personales, familiares, académicos e institucionales. De cada una se muestran algunos estadísticos como: la media, la mediana, mínimo y máximo. En segundo lugar se hace el análisis bivariado a través de tablas de contingencia, pruebas de correlación de Pearson.

La tabla 5 muestra el resumen numérico y descriptivo de las variables objeto de estudio.

Tabla 5 Resumen descriptivo de las variables de estudio del programa de AEA

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NÚMÉRICO
Género	F= Mujer M= Hombre	93 (60 %) 62 (40%)
Nivel educativo de la madre	Básica Primaria Básica Secundaria Técnico Universitario Sin Clasificar = SC	80 (51,61 %) 39 (25,16 %) 5 (3,22%) 8 (5,16 %) 23 (14,38 %)

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NÚMÉRICO
Número de Hermanos	1 = Uno 2= Dos 3 = Tres 4 = Cuatro 5 = Cinco 6 = Seis 7 = Siete Sin Clasificar = SC	12 (7,74%) 59 (38,06 %) 39 (25,16 %) 9 (5,80 %) 5 (3,22 %) 6 (3,87 %) 1 (0.64 %) 24 (15,48 %) Media = 2,69 Min – Max = 1 – 7 Mediana = 2
Edad de Ingreso al programa	En Años	Media = 19,83 Min – Max = 16 - 27 Mediana = 19
Ingreso Familiar al Presentar el Examen de Estado	Alto Medio Bajo Sin Clasificar= SC	2 (1,29) 3 (1,93 %) 148 (95,48 %) 2(1,29)
Trabajaba al presentar el examen de estado	Si No Sin clasificar = SC	9 (5,80 %) 125 (80,64%) 21 (13,54 %)
Vivienda Propia	Si No Sin Clasificar	113 (72,90%) 35(22,58%) 7 (4,51%)
Puntaje en el Examen de Estado	De 1 – 100 puntos	Media = 53,29 Min – Max = 2 - 91 Mediana = 56
Estado del estudiante	Matriculado Desertor Graduado Terminación Académica No matriculado	6 (6,87 %) 72 (46.45 %) 60 (22.4 %) 14 (9.03 %) 1(0,64%)
Número de Materias Perdidas		Media = 9,197 Min – Max = 1 - 34 Mediana = 7
Número de Materias Repetidas		Media = 8,015 Min – Max = 1 - 23 Mediana = 7

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Tiempo Hasta la Deserción	Semestres Académicos (16 semanas)	Media = 4,84 Min – Max = 0 - 16 Mediana = 2
Estado del estudiante Dicotomizado	Desertor No Desertor	72 (46,45 %) 83 (53,55 %)
Tipo de Deserción	Académica No Académica	33 (45,83%%) 39(54,16%)
Motivo de deserción Académica “Perdió Cupo/ Artículo 80”	Literal B	4 (12,1 %)
	Literal C	17 (51,55 %)
	Literal D	7 (21,21%)
	Literal E	5 (15,15 %)
Tipo de Deserción	Precoz	23 (31,94 %)
	Temprana	31(43,05 %)
	Tardía	18 (11,61 %)

Fuente: Autor

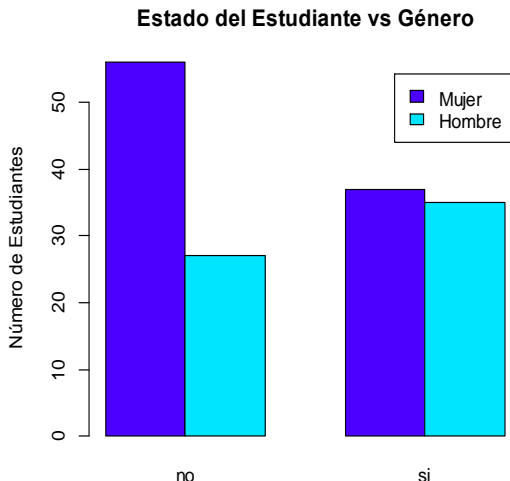
Con los resultados anteriores se puede observar que de los estudiantes seleccionados en la muestra, la mayoría son mujeres (60%) , la edad media al momento de ingresar al programa es de 20 años. El 70.92% tenían vivienda propia y el 80.64% no trabajaban al momento de presentar el examen de Estado.

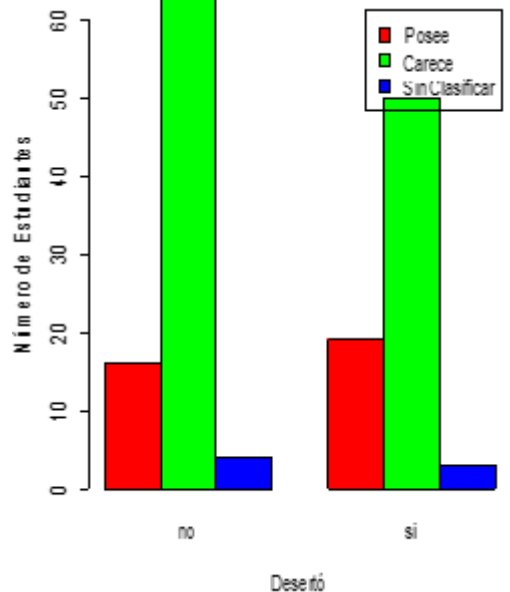
El puntaje promedio al momento de presentar el examen de estado fue de 53.29 en una escala de 0 a 100, el 95,48 % tenía ingresos familiares en nivel bajo al momento de presentar el examen de Estado. Para este conjunto de datos, se tiene que al primer semestre de 2015, ha desertado el 46.45% de los estudiantes de las cohortes entre el primer semestre de 2004 y el segundo semestre de 2009 donde un 243,05% se ubica en deserción temprana.

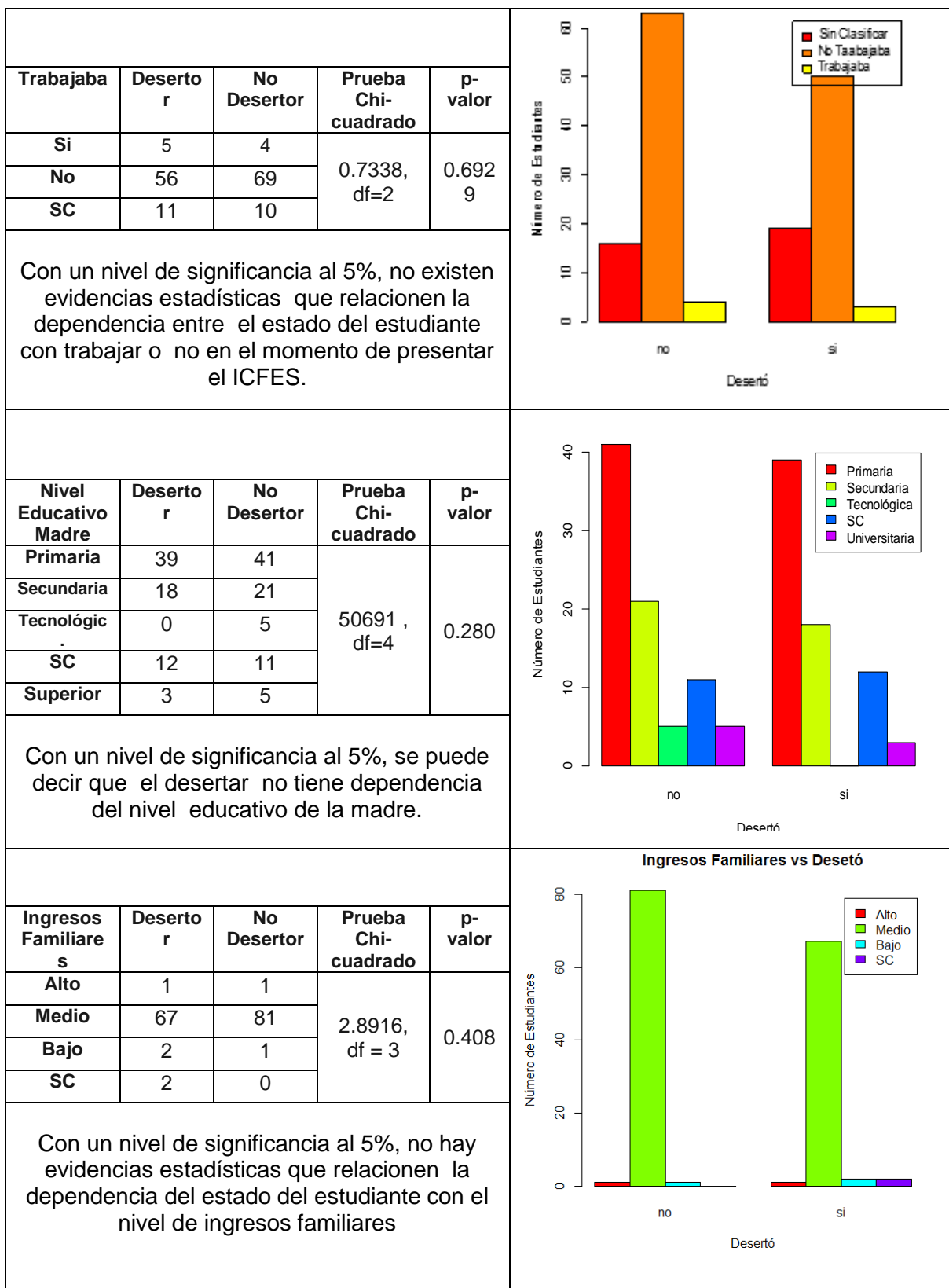
El 38,06 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado, además el 51,61 % de los estudiantes tenían madres con nivel educativo de básica primaria.

Se presenta en la tabla 6 una descripción bivariada de las variables cualitativas de estudio. Se busca determinar si estas tienen dependencia con el estado del estudiante (desertor o no desertor), para lo cual se utilizaron tablas de contingencia y la prueba chi-cuadrado, además se representaron estas en diagramas de barras.

Tabla 6 Contrastes de algunas variables categóricas AEA

					Estado del Estudiante vs Género	
Género	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor		
Hombre	35	27	0.0376, df=4.1543	0.041 53		
Mujer	37	56				
<p>Con un nivel de significancia al 5 %, Se determina que el “Estado del Estudiante” y “Género” son dependientes. Hay una relación entre desertar y género del estudiante.</p>						

					Estado del Estudiante vs Vivienda	
Vivienda	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor		
Posee	19	16	11.20600 df=2	0.571		
Carece	50	63				
SC	3	4				
<p>El valor calculado del p-valor es mayor que 0.05, se rechaza la hipótesis nula de que son dependientes, y por tanto, se dice que el desertar no depende si posee o no vivienda.</p>						



Fuente: Autor

Para comprobar si el tiempo hasta deserción de los estudiantes está correlacionada o no con algunas variables cuantitativas, se realiza un test de correlación de Pearson el cual se presenta en la tabla 7:

Tabla 7 Test de Correlación de Pearson

Variable	Tiempo hasta Desertar		
	Coeficiente de Correlación	Estadístico t	p – valor
Edad de Ingreso al programa	0.011228447	0.0998	0.9208
Número de materias pérdidas	0.3553307	2.895	0.005336
Número de materias repetidas	0.3716149	3.1263	0.002711
Número de hermanos	-0.1068662	-0.7971	0.4288
Puntaje ICFES	-0.1328423	-1.0177	0.3119

Fuente: Autor

Se observa que las variables correlacionas con el tiempo hasta desertar son el número de materias pérdidas y número de materias repetidas, con una correlación leve, a medida que aumenta el tiempo hasta desertar también aumenta el número de materias perdidas y el número de materias repetidas.

4.2. MODELO DE SOBREVIDA PARA DESERCIÓN EN AEA

Como el propósito del estudio es identificar los factores que están relacionados al tiempo de la deserción. En primer lugar, se presentará la aplicación de la metodología de Kaplan Meier para estimar la función de sobrevida, con la cual se puede analizar la evolución de la probabilidad de deserción con su respectivo intervalo de confianza. Posteriormente, se construirá el modelo de regresión de Cox para estimar el efecto de las variables de estudio sobre los tiempos de sobrevivencia a la deserción. Recordemos las características del estudio:

Tabla 8 Consideraciones para el modelo de Sobrevida para AEA

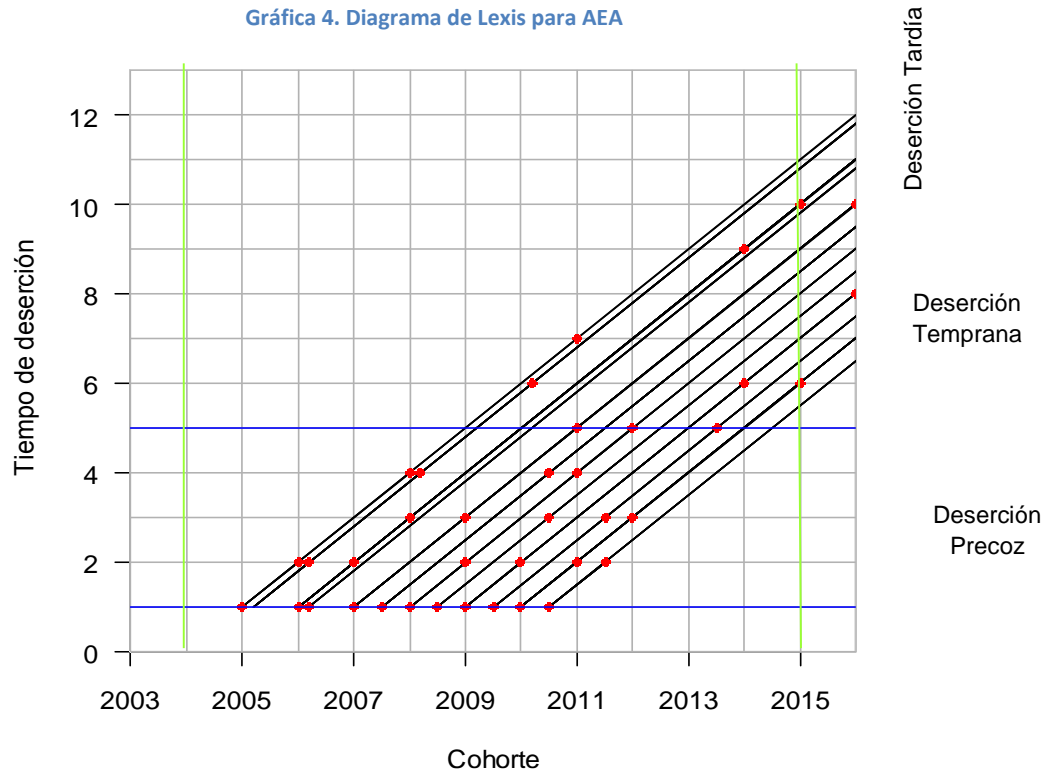
La unidad sobre la cual se registra el evento	Estudiantes seleccionados en la muestra, matriculados en cada programa de administración de empresas agropecuarias en primer semestre.
Evento de interés	Desertar del Programa
Variable respuesta	Tiempo hasta la Deserción: Cuya escala de medición es de razón de tipo discreto, ya que

		se mide en número de semestres cursados hasta que presenta el fenómeno de estudio.
Tiempo	Inicial del Estudio	I Semestre de 2004
	Origen del Evento	Cohorte del Estudiante (primera matricula del estudiante por programa)
	Final del Estudio	I Semestre de 2015
La escala de medida del tiempo hasta el evento		Semestres Calendario(16 Semanas)
Censura	Tipo	Tipo I (a derecha)
	Tiempo	13 semestres por cohorte

Fuente: Autor

En seguida se presenta la representación de los estudiantes del estudio mediante el diagrama de Lexis, cuya sintaxis R aparece en el Anexo E. Diagrama que refleja el tiempo calendario en el eje horizontal y la longitud del tiempo de vida, representada por una línea a 45°. El tiempo que un individuo pasa en el estudio es representado por la altura del rayo en el eje vertical.

Gráfica 4. Diagrama de Lexis para AEA



Fuente: Autor

En la gráfica 4 evidencia que los estudiantes bajo estudio no tienen el mismo tiempo de origen, se señala con los puntos rojos aquellos que desertaron y el tiempo en que sucedió (expresado en semestres). Nótese que la mayoría de estudiantes desertan precozmente, seguido de la deserción temprana, unos pocos se registran con deserción tardía.

Ahora se presenta el análisis del tiempo hasta la deserción, teniendo en cuenta toda la información disponible, es decir tanto los datos censurados como los no censurados, razón por la cual, las probabilidades de sobrevivencia en cada intervalo así como la función de sobrevivencia se calculan con el estimador de Kaplan Meier.

4.2.1 Función de Sobrevida

El estimador de Kaplan-Meier para la función de sobrevivencia es obtenido en el paquete estadístico R mediante la función `survfit`, en “Anexo F” se describe la sintaxis y en seguida se muestra la salida de R en la tabla 9, donde: *time* representa el tiempo en semestres, *n.risk* indica la cardinalidad del conjunto en riesgo o el número de individuos que continua en estudio al tiempo correspondiente, *n.event* corresponde al número de fallas que se presentan entre cada tiempo, *survival* indica el valor que toma la función de sobrevivencia estimada por el método Kaplan-Meier en el tiempo correspondiente, *std.err* corresponde al error estándar estimado para la función de supervivencia en el tiempo respectivo y finalmente *lower95 %CI* y *upper95 %CI* denotan el intervalo de confianza para la función de sobrevivencia en cada tiempo. En la gráfica 5 se pueden observar las probabilidades de sobrevivencia en cada semestre, así:

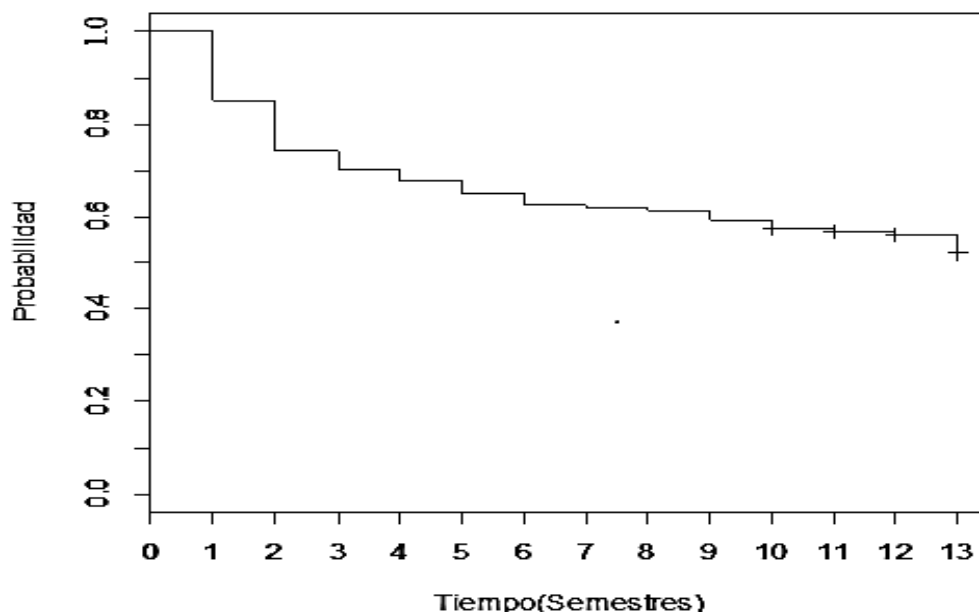
Tabla 9 Resumen de K-M para AEA

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	155	23	0.852	0.0286	0.797	0.909
2	132	17	0.742	0.0351	0.676	0.814
3	115	6	0.703	0.0367	0.635	0.779
4	109	4	0.677	0.0375	0.608	0.755
5	105	4	0.652	0.0383	0.581	0.731
6	101	4	0.626	0.0389	0.554	0.707
7	97	1	0.619	0.0390	0.547	0.701
8	96	1	0.613	0.0391	0.541	0.695
9	95	3	0.594	0.0395	0.521	0.676
10	92	3	0.574	0.0397	0.501	0.658
11	86	1	0.568	0.0398	0.495	0.651
12	78	1	0.560	0.0400	0.487	0.644
13	57	4	0.521	0.0417	0.445	0.609

Fuente: Autor

La probabilidad de continuar con el programa va disminuyendo a medida que el estudiante avanza durante el programa. Nótese que es más rápido el decrecimiento en los primeros seis semestres y tiende a estabilizarse gradualmente a partir del 6 semestre donde está alrededor de 62,6%. Por ejemplo la probabilidad de continuar con el programa hasta el primer semestre es del 85,2% y baja a 74,2% en el segundo semestre, luego en el tercer semestre esta posibilidad es del 70,3%, llegando así con un 52,1% al semestre 13.

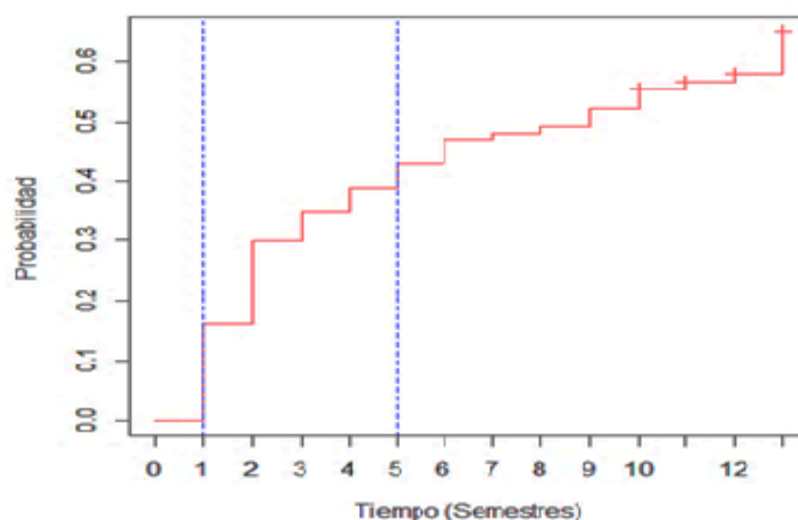
Gráfica 5 Función de Sobrevida de AEA



Fuente: Autor

La función de riesgo, también conocida como la tasa instantánea de mortalidad, describe la forma en que cambia la tasa instantánea de la deserción al paso del tiempo, y la función de riesgo acumulado, permite tener información del comportamiento del riesgo a lo largo del tiempo. A continuación se presenta la correspondiente al programa de AEA, para efectos de interpretación se indica en la gráfica 6 el semestre 1 (deserción precoz) y semestre 5 (deserción temprana).

Gráfica 6 Función de riesgo acumulada de AEA



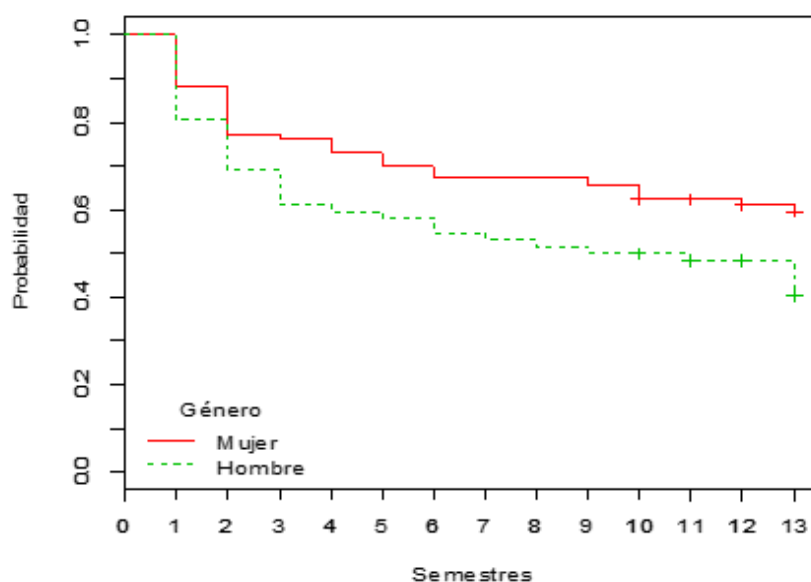
Fuente: Autor

Con base en la función de riesgo acumulada se puede afirmar que la probabilidad que un estudiante deserte justo al terminar el primer semestre es del 16 %, es decir que hay una probabilidad baja que deserte precozmente del Programa. De igual manera se observa que la función de riesgo presenta los mayores cambios en la deserción temprana, por ejemplo, se puede apreciar, que se tiene un punto crítico en segundo semestre donde la probabilidad de que habiendo llegado a primero deserte en el segundo semestre es del 14 %. Además la posibilidad de que llegado al segundo deserte en el tercero es alrededor de 29,8%.

Teniendo en cuenta que el fenómeno de la deserción no se está estudiando de manera aislada, se hace necesario un análisis de la función de sobrevivencia separado por poblaciones de estudio, lo cual se presenta a continuación. En primer lugar, la función de sobrevivencia por género: Al comparar las curvas de sobrevivencia se obtiene un valor de $\chi^2 = 4.5$ con un grado de libertad y un p valor de 0.0345, estadística que permite afirmar que se rechaza la hipótesis nula de igualdad de las

curvas. Por lo tanto se puede afirmar que la sobrevivencia (a la deserción) de hombres y mujeres no tiene el mismo comportamiento, tal como se aprecia en la gráfica 7. Las mujeres tienen más probabilidad de sobrevivir en relación con los hombres. A partir del primer semestre se vuelve diferente este comportamiento y se extiende hasta el semestre trece donde la probabilidad de no desertar para el género masculino se encuentra 19,4% por debajo en relación con las mujeres que es de 59,7%.

Gráfica 7 Función de Sobrevida por Género de AEA



Fuente: Autor

Tabla 10 Estimador K-M por género

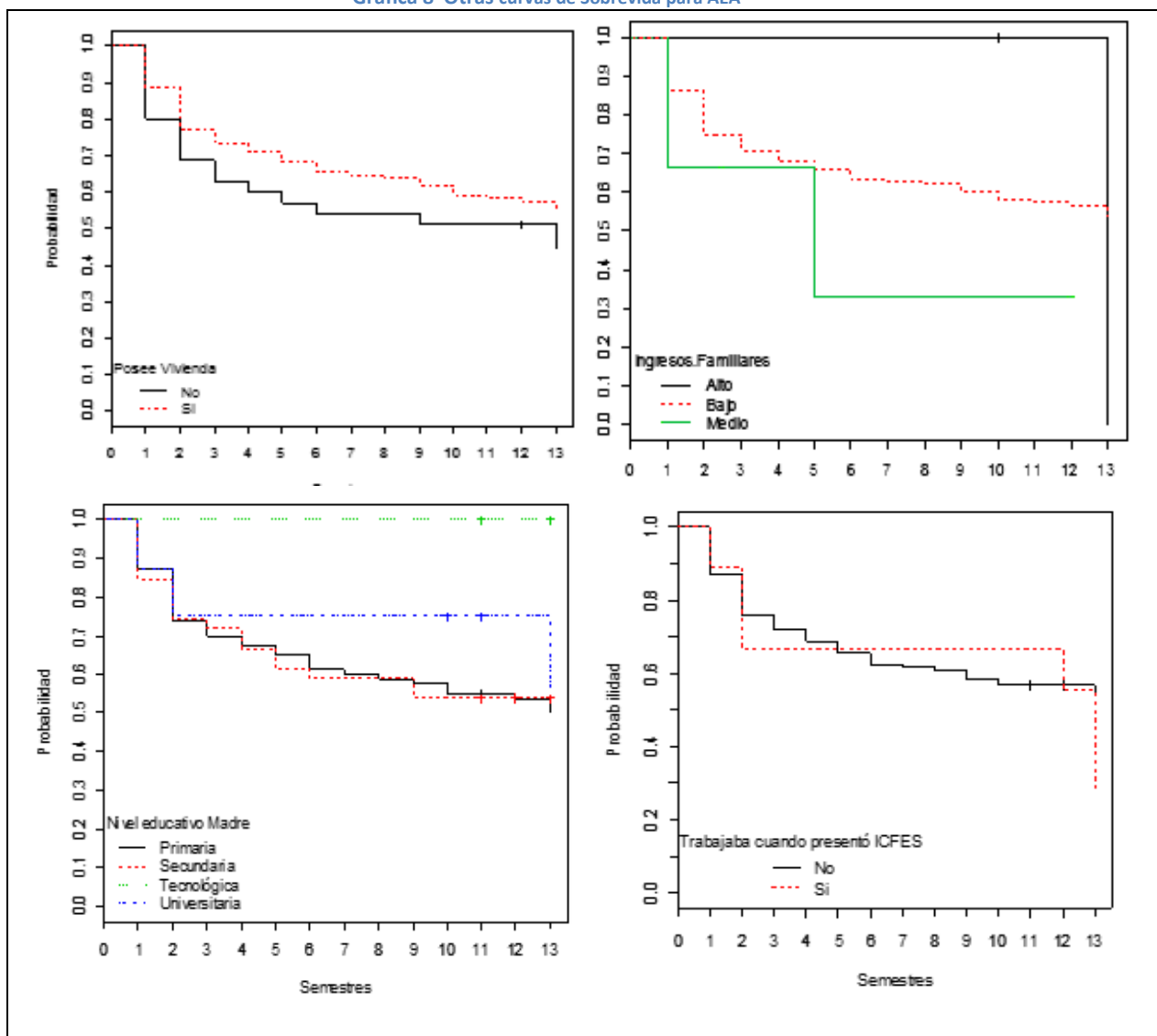
Género=Femenino							Género=Masculino						
					95% IC							95% IC	
Time	n.risk	n.event	survival	std.err	lower	upper	Time	n.risk	n.event	Survival	std.err	lower	upper
1	93	11	0.882	0.0335	0.818	0.950	1	62	12	0.806	0.0502	0.714	0.911
2	82	10	0.774	0.0434	0.694	0.864	2	50	7	0.694	0.0585	0.588	0.818
3	72	1	0.763	0.0441	0.682	0.855	3	43	5	0.613	0.0619	0.503	0.747
4	71	3	0.731	0.0460	0.646	0.827	4	38	1	0.597	0.0623	0.486	0.732
5	68	3	0.699	0.0476	0.612	0.799	5	37	1	0.581	0.0627	0.470	0.717
6	65	2	0.677	0.0485	0.589	0.779	6	36	2	0.548	0.0632	0.438	0.687
9	63	2	0.656	0.0493	0.566	0.760	7	34	1	0.532	0.0634	0.421	0.672
10	61	3	0.624	0.0502	0.533	0.730	8	33	1	0.516	0.0635	0.406	0.657
12	53	1	0.612	0.0506	0.520	0.720	9	32	1	0.500	0.0635	0.390	0.641
13	40	1	0.597	0.0516	0.503	0.707	11	30	1	0.483	0.0635	0.374	0.625
							13	18	3	0.403	0.0679	0.289	0.560

Fuente: Autor

Se efectuaron las pruebas de LongRank para juzgar la igualdad de curvas de sobrevivencia para las variables que se muestran en el Gráfico 8, encontrándose que para las diferentes variables no rechaza la hipótesis nula de igualdad de estas, como lo resume a continuación:

Prueba de LongRank			
Variable	Chisq	Grados de Libertad	P valor
Nivel.educativo.madre	3.7	3	0.296
Tipo.de.Vivienda	1.2	1	0.277
Trabajaba.presento.ICFES	0.8	2	0.668
Ingresos.Familiares	0.9	2	0.634

Gráfica 8 Otras curvas de Sobrevida para AEA



Fuente: Autor

4.2.2. Factores asociados al riesgo de deserción

En el apartado anterior se hizo un análisis de la situación de sobrevivencia de los estudiantes del programa de AEA, en esta sección se pretende llegar a explicar los factores relacionados con el riesgo de deserción en los estudiantes mediante el modelo de regresión de Cox. Inicialmente se presenta el modelo completo con todas las variables, luego para la obtención del mejor conjunto de variables explicativas del modelo se tiene en cuenta el Criterio de Información de Akaike (AIC), con el método de selección hacia adelante. Posteriormente se calcula la curva de Sobrevivencia. Las variables a considerar son el género, edad de ingreso al programa, nivel de ingresos familiares, nivel educativo de la madre, puntaje estandarizado del ICFES, Estado de empleo cuando presentó el ICFES, y tenencia de vivienda.

El modelo de regresión de Cox inicial con todas las covariables, se presenta continuación:

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	3.058e-03	1.003e+00	7.044e-02	0.043	0.965
Trabajaba.cuando.presento.ICFES[T.Si]	1.315e-01	1.141e+00	6.131e-01	0.214	0.830
Género[T.M]	3.497e-01	1.419e+00	3.110e-01	1.124	0.261
Ingresos.Familiares[T.Bajo]	-8.328e-01	4.348e-01	1.457e+00	-0.572	0.568
Ingresos.Familiares[T.Medio]	3.024e+00	2.057e+01	1.501e+00	2.015	0.044 *
Nivel.educativo.madre[T.Secundaria]	-4.414e-02	9.568e-01	3.981e-01	-0.111	0.912
Nivel.educativo.madre[T.Tecnológica]	-1.739e+01	2.795e-08	3.564e+03	-0.005	0.996
Nivel.educativo.madre[T.Universitaria]	-5.875e-01	5.557e-01	1.069e+00	-0.549	0.583
No..De.hermanos	-1.625e-02	9.839e-01	1.479e-01	-0.110	0.912
Puntaje.estandarizado.ICFES	2.864e-03	1.003e+00	8.033e-03	0.357	0.721
Tipo.de.Vivienda[T.Posee]	-4.531e-01	6.357e-01	3.279e-01	-1.382	0.167

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Seguidamente se procede a seleccionar el modelo más parsimonioso, utilizando el método hacia adelante y usando como criterio el AIC (en el Anexo D se muestra en detalle esta selección), encontrándose que el modelo óptimo queda determinado por el género e ingresos familiares. Sin embargo desde el marco conceptual del problema se decide considerar en el modelo la edad de Ingreso al programa, ya

que revisando la literatura se tiene que en los estudios realizados por Giovanoly (2002) y Barrera (2009) es relevante a la hora de explicar el tiempo hasta que un estudiante deserta.

La salidas en **R** que se presentarán dan cuenta de información acerca de las pruebas locales para verificar que cada coeficiente es significativamente distinto de cero, estos coeficientes estimados, se consideran significativos cuando el cociente $z = \left| \frac{coef}{se(coef)} \right|$ en valor absoluto, es superior a 2, ya que para muestras grandes este cociente se distribuye según una ley Normal (prueba de Wald).

Las columnas proporcionan información para cada covariable como sigue:

coef : El valor del coeficiente de regresión estimado.

exp(coef) : La función exponencial evaluada en el coeficiente de regresión estimado, (corresponde al riesgo relativo).

se(coef): el error estándar del coeficiente de regresión estimado.

z : Corresponde a la estadística de prueba, obtenida dividiendo el valor del coeficiente de regresión estimado entre el error estándar estimado.

p: El p-value que corresponde a dos veces el área acumulada a la derecha del cuantil *z* en una distribución normal con media cero y varianza uno.

También se presenta información acerca de los coeficientes de regresión estimados, tal como:

exp(coef) : La función exponencial evaluada en el coeficiente de regresión estimado.

exp(-coef) : La función exponencial evaluada en, menos el coeficiente de regresión estimado(útil para la interpretación cuando el (*coef*) es negativo que indica factores de protección).

lower .95 : El límite inferior de un intervalo del 95 % de confianza para el coeficiente de regresión estimado.

upper .95 : El límite superior de un intervalo del 95 % de confianza para el coeficiente de regresión estimado.

Retomando, la estimación de los parámetros del modelo de Cox seleccionado, corresponde a:

```
n= 104, number of events= 50
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-0.001483	0.998519	0.061913	-0.024	0.9809
Género[T.M]	0.499915	1.648581	0.308053	1.623	0.1046
Ingresos.Familiares[T.Bajo]	-0.169640	0.843968	1.033762	-0.164	0.8697
Ingresos.Familiares[T.Medio]	3.089620	21.968723	1.489964	2.074	0.0381 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Edad.de.ingreso.al.programa	0.9985	1.00148	0.8844	1.127
Género[T.M]	1.6486	0.60658	0.9014	3.015
Ingresos.Familiares[T.Bajo]	0.8440	1.18488	0.1113	6.401
Ingresos.Familiares[T.Medio]	21.9687	0.04552	1.1845	407.440

Se observa que las características observadas en los estudiantes, se tiene que tan sólo el nivel de ingresos familiares es significativa a la hora de explicar el riesgo de deserción.

En cuanto a la edad de ingreso al programa, la estimación puntual es negativa e indica que por cada año más de edad de ingreso al programa disminuye el riesgo de desertar, dejando las otras variables constantes, si se comparan dos estudiantes, aquel estudiante mayor en un año de edad tiene 1.001 veces más probabilidad de desertar. Se observa que el p-valor es 0.9809 (mayor que 0.05), por lo que la relación entre la edad y la sobrevivencia estimada no es estadísticamente significativa.

En cuanto al nivel de ingresos familiares [Medio], indica que al tener este nivel de ingresos se tiene 21.97 veces mayor probabilidad de desertar que los individuos que registraron tener ingresos altos.

Por su parte los Hombres (M) tienen un riesgo de 1.65 mayor de desertar respecto de las mujeres, hecho que es consistente con otros estudios, donde afirman que las mujeres poseen mayor responsabilidad y perseverancia en sus proyectos

educativos según Giovanoly (2002). Además el p-valor es mayor que 0.05, hecho que corrobora que la edad y la sobrevivencia no son estadísticamente significativas.

Los estudiantes que reportaron ingresos familiares en nivel bajo disminuye el riesgo de desertar en 1.19 respecto a los estudiantes que al momento de presentar el ICFES registraban tener ingresos familiares altos, esta variable no es significativa al momento de explicar la deserción.

En resumen se puede decir que los factores que aumentan el riesgo de deserción son reportar un nivel ingresos familiares Medio y ser Hombres, por su parte se disminuye el riesgo de presentar el evento a aquellos que tenían ingresos familiares en nivel bajo y registran una mayor edad de ingreso al programa.

Al juzgar la hipótesis de significancia del modelo se rechaza la hipótesis para las pruebas de Wald y Score, tal como se aprecia en la tabla 11. Las variables seleccionadas para el estudio tan sólo explican el 6.6 % de la variabilidad en el tiempo de deserción. Aspecto que indica que al analizar la deserción se deben considerar otras características como los personales, familiares, institucionales, socioeconómicos.

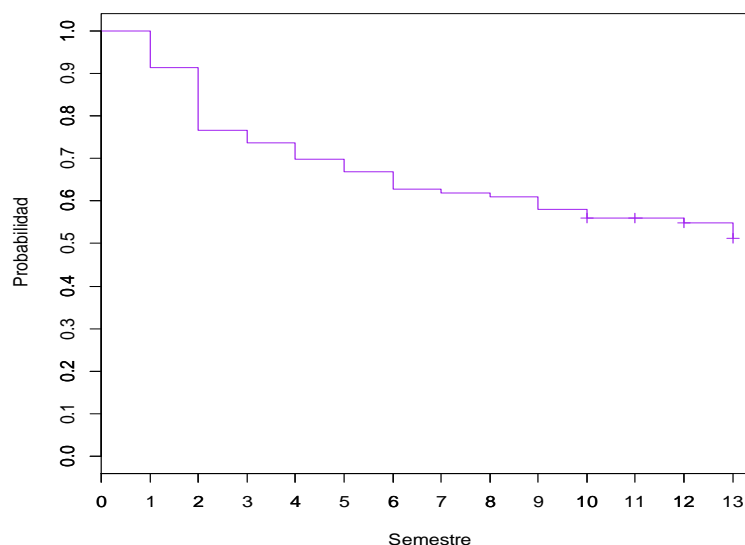
Tabla 11 Pruebas de Verosimilitud, Wald y Score.

Rsquare= 0.066 (max possible= 0.985)			
Likelihood ratio test=	7.1	on 4 df,	p=0.1306
Wald test	= 11.05	on 4 df,	p=0.026
Score (logrank) test =	19.65	on 4 df,	p=0.0005843

Fuente: Autor

A continuación en la gráfica 9 y de manera más detallada en la tabla 12 se ilustra la función de sobrevivencia ajustada al modelo de regresión de Cox con las variables predictoras género, ingresos familiares y Edad de Ingreso al Programa.

Gráfica 9 Función de Sobrevida para el modelo de AEA



Fuente: Autor

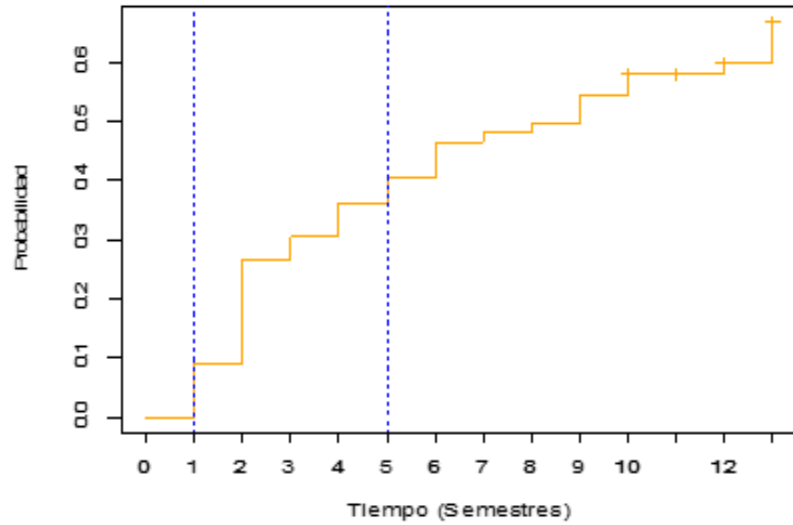
Tabla 12 Función de sobrevida del modelo en AEA

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	104	10	0.912	0.0277	0.860	0.969
2	94	15	0.766	0.0419	0.688	0.853
3	79	3	0.736	0.0437	0.656	0.827
4	76	4	0.697	0.0456	0.613	0.793
5	72	3	0.668	0.0468	0.582	0.766
6	69	4	0.629	0.0481	0.541	0.730
7	65	1	0.619	0.0484	0.531	0.721
8	64	1	0.609	0.0486	0.521	0.712
9	63	3	0.579	0.0493	0.490	0.684
10	60	2	0.559	0.0496	0.470	0.665
12	56	1	0.549	0.0497	0.460	0.656
13	47	3	0.513	0.0507	0.422	0.623

Fuente: Autor

Se observa en la curva de riesgo acumulado en la gráfica 10, que la probabilidad de que un estudiante deserte justo al terminar el primer semestre es del 9,2%, además se tiene un punto crítico en segundo semestre donde la probabilidad de que habiendo llegado a primero deserte en el segundo semestre es del 26,6% y de que llegando al segundo deserte en el tercer semestre es del 30%. Donde el ajuste que se obtiene de las probabilidades corresponde a la inclusión en su cálculo de la edad de ingreso, el género y el nivel de ingreso familiar.

Gráfica 10 Función de riesgo acumulada del modelo de Cox para AEA



Fuente: Autor

4.2.4. Pronóstico a partir del modelo

En la sección 2.4.5 se mencionó que la función de riesgo inicial $h_0(t)$ depende del tiempo, mientras que $\exp(X\beta^T)$ depende únicamente de las covariables. Sin embargo, en el modelo de riesgos proporcionales de Cox se debe tener en cuenta que los riesgos para dos conjuntos diferentes de valores de los covariables conservan la misma proporción a lo largo del tiempo, de ahí su denominación.

Con base en los resultados obtenidos de los factores asociados al riesgo de deserción en AEA (sección 4.2.2) se describen a continuación:

$$\begin{aligned}
 h_i(t, X, \beta) = & h_0(t) \exp -0.0011483(Edadingreso) + 0.499915Géner0[T. M]) \\
 & + (-0.169640(Ing. Famili[T. Bajo])) \\
 & + (3.0896220(Ing. Famili[T. Medio])) \quad (4.24)
 \end{aligned}$$

El exponente del modelo de Cox, que particularmente para el estudio está contenido en la ecuación (4.24), se denomina *índice pronóstico* (P_I), (Domnéch, 1992). De manera que la tasa de riesgo se puede expresar a través de este indicador en lugar

utilizar los valores del estudiante para cada variable. En la práctica es mejor usar el *índice de pronóstico centrado* (P_{IC}), que se obtiene centrando las variables predictoras:

$$P_{IC} = \beta_1(X_1 - \bar{X}_1) + \dots + \beta_p(X_p - \bar{X}_p)$$

Que para este caso corresponde a:

$$\begin{aligned} P_{IC} = & (-0.0011483((X_1 - 19.83) + 0.499915Géner0[T.M]) \\ & + (-0.169640(Ing.Famili[T.Bajo])) \\ & + (3.0896220(Ing.Famili[T.Medio]))) \end{aligned}$$

Un índice pronóstico centrado que sea igual a cero para el programa de AEA, corresponde entonces a un estudiante cuyas características son el ser mujer, con edad promedio de 19.03 años y de estrato alto y que se denominará de ahora en adelante “estudiante base”. La diferencia entre los índices pronósticos de dos estudiantes permite estimar su riesgo relativo (RR), veamos:

$$RR = \frac{h_i(t, X_i, \beta)}{h_j(t, X_j, \beta)} = \frac{h_0(t) \exp(P_{IC\text{individuo}i})}{h_0(t) \exp(P_{IC\text{individuo}j})} = \exp(P_{IC\text{individuo}i} - P_{IC\text{individuo}j})$$

Teniendo en cuenta lo anterior, a continuación se presenta el RR respecto al “estudiante base” para cinco estudiantes, seleccionados aleatoriamente, y que se encontraban cursando el primer semestre del Programa, es decir, estudiantes que ingresaron en el I semestre de 2015.

Tabla 13 Pronóstico del modelo Individual en AEA

N°	Estudiante			P_{IC}	RR respecto al estudiante base
	Género	Edad.Ingr	Ingre.Fam		
1	F	18	Bajo	-0.1717	1.187371
2	F	17	Medio	3.0929	22.0425
3	M	18	Bajo	0.3323	1.3942
4	M	19	Medio	3.5904	36.2517
5	M	17	Bajo	0.3335	1.3959

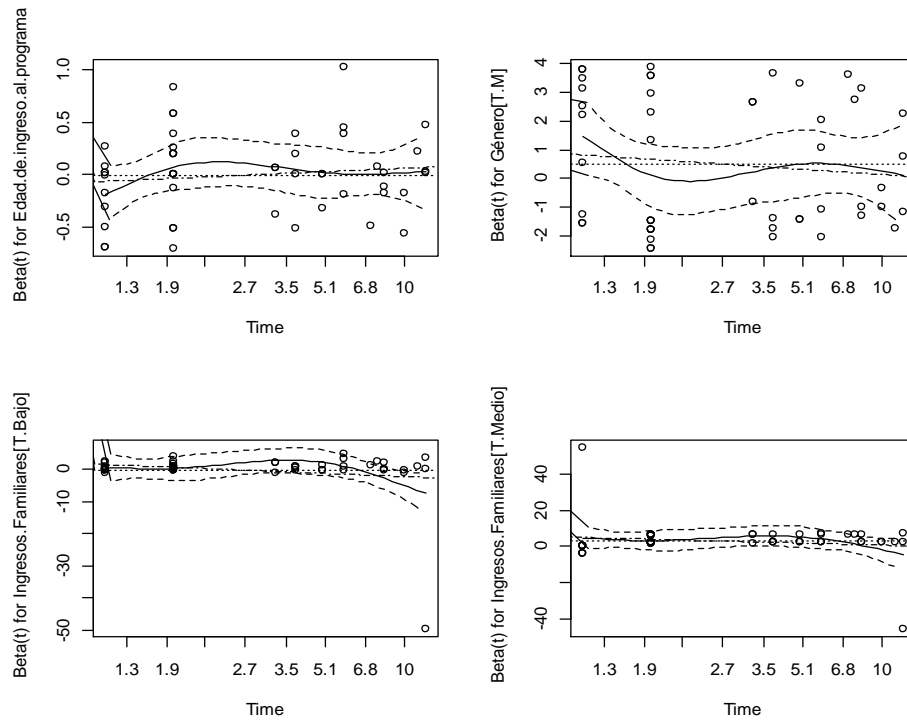
Lo anterior significa que respecto al estudiante base el riesgo de deserción para el estudiante 1 es menor. Es decir que el riesgo de deserción para el estudiante base es 1.187371 veces mayor que el de un estudiante mujer con 18 años de edad e ingreso familiar bajo. Ahora bien, respecto al estudiante base, un estudiante mujer con 17 años de edad y estrato medio tiene 22.0425 veces mayor riesgo de deserción que el estudiante base. De manera similar se efectúa la interpretación para los estudiantes 3,4 y 5.

4.2.3. Evaluación del modelo

Al verificar el supuesto de riesgos proporcionales del ajuste del modelo de regresión para AEA en donde se obtiene que:

	rho	chisq	p
Edad.de.ingreso.al.programa	0.117	0.511	0.475
Género[T.M]	-0.111	0.659	0.417
Ingresos.Familiares[T.Bajo]	-0.198	1.908	0.167
Ingresos.Familiares[T.Medio]	-0.152	1.132	0.287
GLOBAL		2.902	0.574

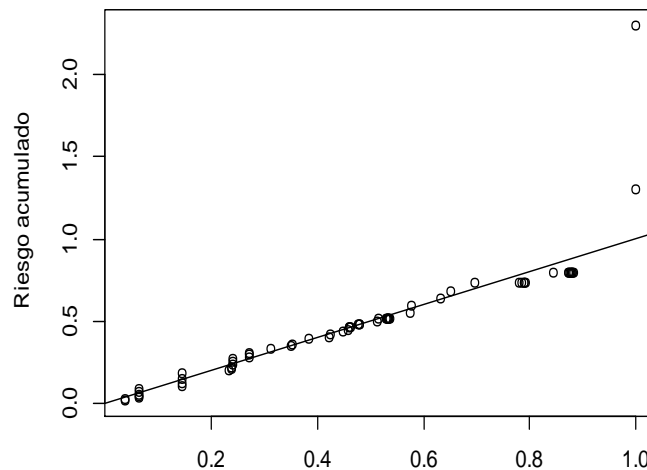
Por lo que, existen evidencias de que no se está violando el supuesto de riesgos proporcionales de manera global, ni por ninguna covariable. Además este supuesto se aborda de manera gráfica a través de los residuos escalados de Schoenfeld, mencionadas en la sección 2.4.4 respecto a cada covariable. La gráfica 11 evidencia que ningún coeficiente es dependiente del tiempo.



Gráfica 11 Residuos de Schoenfeld

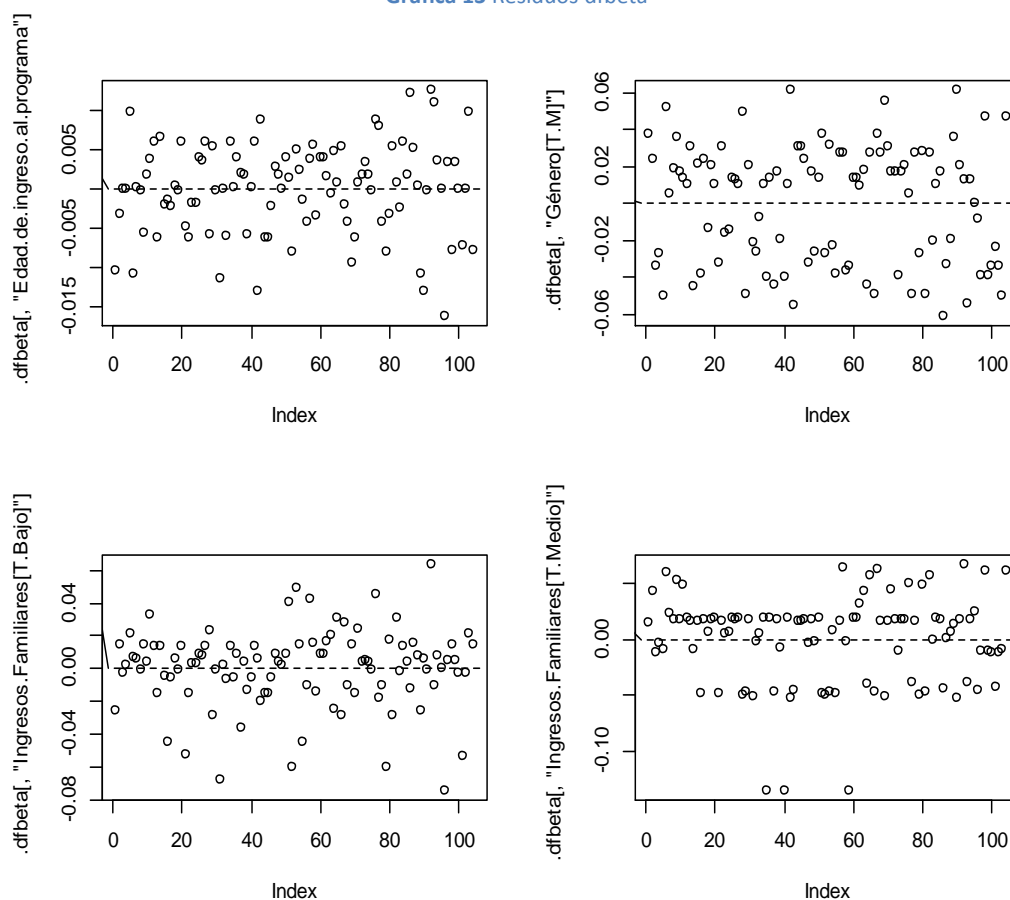
Comprobación de la hipótesis global del modelo mediante los residuos de Cox-Snell: Si el modelo es correcto y la estimación de los β son cercanas a los valores reales, entonces este estimador debería describir aproximadamente una recta de pendiente igual a la unidad. La gráfica 12 Indica que, en general, este modelo se ajusta a los datos.

Gráfica 12 Residuos de Cox-Snell para el modelo de Deserción en AEA



Residuos dfbeta: Con este tipo de residuo, se determina la influencia de cada observación en el modelo. De manera que, para cada covariable, se ha representado la observación (en orden de tiempo de falla registrado) por el cambio de escala aproximada del coeficiente, es decir, dividiendo por el error estándar, después de la eliminación de la observación del modelo.

Gráfica 13 Residuos dfbeta



Fuente: Autor

En este caso ninguna de las observaciones ejerce un cambio mayor a aproximadamente un 10% del error estándar. Si se comparan con el valor coeficiente de regresión estimado y su error estándar podemos concluir que ninguna de las observaciones es influyente individualmente.

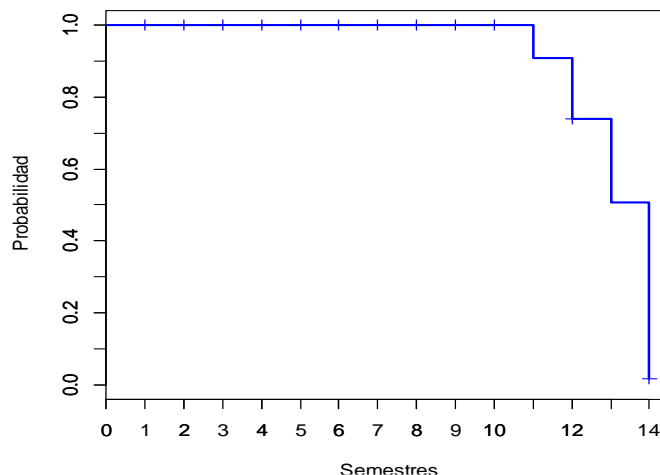
4.3. MODELO PARA GRADUACIÓN EN AEA

La duración teórica de la carrera universitaria es de cinco años. Sin embargo, los alumnos que no desertan prolongan sus estudios por un mayor tiempo. Este apartado se concentra en una pregunta: ¿Qué factores “favorecen” a la finalización exitosa del programa de AEA dado el tiempo que lleva el alumno estudiando?. Nótese que se sigue presentando censura tipo I a derecha, y el tiempo de censura se define a partir del *cuartil 3* del tiempo de permanencia de los estudiantes en el programa que corresponde a 14 semestres.

4.3.1. Función de Supervivencia

En la gráfica 14 se observa que la probabilidad de los estudiantes que llegan a 10 semestres y no se han graduado es alta (algo que era de esperarse), además la probabilidad de llegar y de no graduarse en el semestre 12 es de 73.95%, mientras que la de llegar al semestre 13 sin graduarse es del 50.73%.

Gráfica 14 Función de Supervivencia para la Graduación en AEA



Fuente: Autor

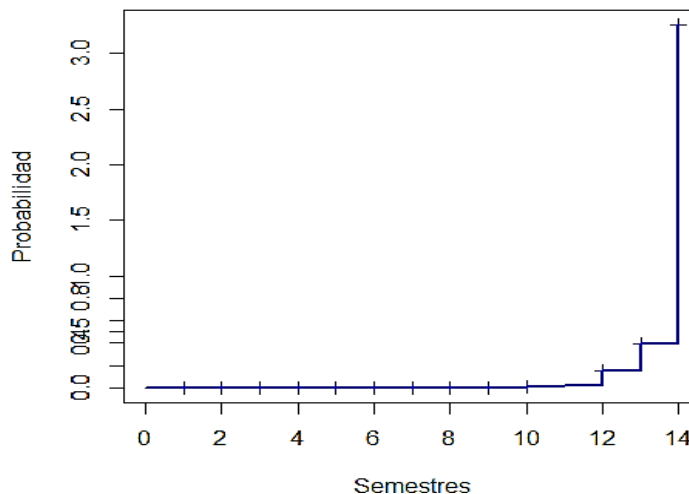
Tabla 14 Estimador de K-M Graduación en AEA

					95% CI	
Time	n.risk	n.event	survival	std.err	Lower	Upper
11	119	11	0.9076	0.0266	0.85699	0.9611
12	108	20	0.7395	0.0402	0.66470	0.8227
13	86	27	0.5073	0.0462	0.42445	0.6064
14	59	57	0.0172	0.0121	0.00435	0.0679

Fuente: Autor

Se muestra la función de riesgo acumulada en la gráfica 15, la cual indica que la probabilidad de terminar el undécimo semestre y graduarse es 10%, en el semestre número doce la probabilidad de graduarse es del 30.17% y por último en el semestre 14 la probabilidad de graduarse al terminarlo es muy alta respecto a los anteriores semestres.

Gráfica 15 Función de riesgo acumulada para graduación



Fuente: Autor

4.3.2 Factores relacionados al aumento de la graduación

En seguida se obtiene el modelo de regresión de Cox con todas las covariables disponibles, adicionado a ellas dos de tipo académico, número de asignaturas pérdidas, y número de materias repetidas.

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-5.090e-02	9.504e-01	8.274e-02	-0.615	0.5384
Trabajaba.cuando.presento.ICFES[T.Si]	5.968e-01	1.816e+00	7.334e-01	0.814	0.4158
Género[T.M]	3.180e-01	1.374e+00	3.613e-01	0.880	0.3788
Ingresos.Familiares[T.Bajo]	1.738e+01	3.537e+07	5.055e+03	0.003	0.9973
Ingresos.Familiares[T.Medio]	NA	NA	0.000e+00	NA	NA
Número.de.materias.repetidas.	-1.324e-01	8.760e-01	7.266e-02	-1.822	0.0684
Nivel.educativo.madre[T.Secundaria]	-3.597e-01	6.979e-01	4.912e-01	-0.732	0.4640
Nivel.educativo.madre[T.Tecnológica]	-1.027e-01	9.024e-01	6.584e-01	-0.156	0.8761
Nivel.educativo.madre[T.Universitaria]	1.585e-02	1.016e+00	7.532e-01	0.021	0.9832
No..De.hermanos	1.159e-02	1.012e+00	1.462e-01	0.079	0.9368
Puntaje.estandarizado.ICFES	-1.815e-02	9.820e-01	9.507e-03	-1.909	0.0563
Tipo.de.Vivienda[T.Posee]	-1.248e-01	8.827e-01	3.730e-01	-0.335	0.7379
X.Materias.perdidas.	1.301e-02	1.013e+00	4.955e-02	0.263	0.7929

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Y siguiendo el criterio de selección hacia adelante, se obtiene el modelo óptimo con las covariables número de materias pérdidas, puntaje ICFES, ingresos familiares, mostrando la siguiente salida de **R** del modelo de Cox:

	coef	exp(coef)	se(coef)	z	Pr(> z)
Puntaje.estandarizado.ICFES	-1.199e-02	9.881e-01	6.398e-03	-1.874	0.06093
Ingresos.Familiares[T.Bajo]	1.709e+01	2.653e+07	3.677e+03	0.005	0.99629
Ingresos.Familiares[T.Medio]	1.861e+01	1.208e+08	3.677e+03	0.005	0.99596
Número.de.materias.repetidas.	-8.254e-02	9.208e-01	2.626e-02	-3.143	0.00167 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Puntaje.estandarizado.ICFES	9.881e-01	1.012e+00	0.9758	1.0006
Ingresos.Familiares[T.Bajo]	2.653e+07	3.769e-08	0.0000	Inf
Ingresos.Familiares[T.Medio]	1.208e+08	8.277e-09	0.0000	Inf
Número.de.materias.repetidas.	9.208e-01	1.086e+00	0.8746	0.9694

En cuanto a las variables incluidas dentro del modelo, el número de materias repetidas y el puntaje ICFES son significativas a la hora de explicar la graduación.

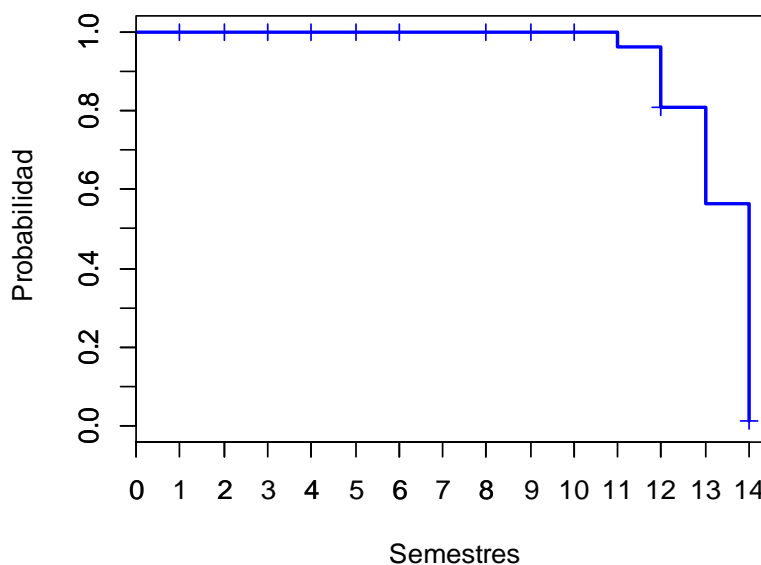
Por su parte los estudiantes que ingresan con un puntaje alto del ICFES, tienen un riesgo relativo de 0.9881, con un intervalo de confianza al 95% (0.9758, 1.006). Lo que significa que si se comparan dos estudiantes, manteniendo las demás variables constantes, para aquel estudiante que tenga un punto más en la prueba, se disminuye en 1.012 la probabilidad de graduarse.

El registrar ingresos familiares bajos aumenta en 17.09 veces la probabilidad de graduarse respecto de los estudiantes con ingresos familiares altos. Además el tener ingresos familiares en nivel medio también aumenta la probabilidad manteniendo las demás variables constantes de graduarse en 18.61 veces más que tenían ingresos altos.

Los resultados indican que riesgo relativo al número de materias repetidas es de 0.9081, con un intervalo de confianza al 95%(0.8746, 0.9604). Como la estimación puntual es inferior a 1, señala que si se comparan dos estudiantes, manteniendo las demás variables constantes, aquel estudiante que haya repetido una materia más que otro, disminuye la probabilidad en 1.086 veces de graduarse.

A continuación en la gráfica 16 se presenta la función de sobrevivida del modelo anterior, se aprecia cómo es el comportamiento de la graduación de los estudiantes de AEA, vale la pena resaltar que habiendo llegado un estudiante a 10 semestre la probabilidad de que no se haya graduado es alta, de llegar al semestre 11 donde la probabilidad es del 96%, de llegar al 12 y no graduarse corresponde a 56%.

Gráfica 16 Función de sobrevivida para el modelo de Graduación en AEA.



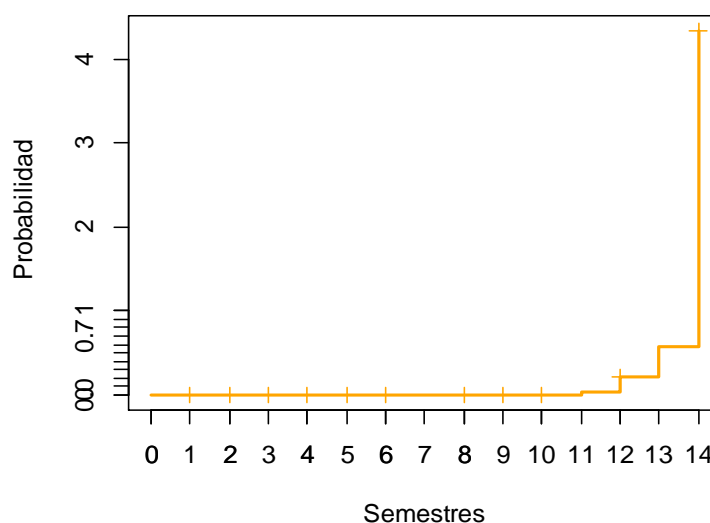
Fuente: Autor

Tabla 15 Función de sobrevivencia del modelo para Graduación.

Time	n.risk	n.event	Survival	std.err	lower95% CI	upper 95%CI
11	79	3	0.9620	0.0215	0.92079	1.000
12	76	12	0.8101	0.0441	0.72810	0.9014
13	62	19	0.5619	0.0564	0.46144	0.6841
14	43	42	0.0131	0.0130	0.00186	0.0916

Fuente: Autor

Gráfica 17 Función de Riesgo del modelo de Graduación



Fuente: Autor

Para la bondad de ajuste del modelo, se aprecia en la tabla 16 un p-valor significativamente pequeño, lo cual es evidencia de que las covariables seleccionadas explicativas tienen coeficientes distintos de cero en el modelo de riesgos proporcionales, además el modelo explica el 16.6 % de la variabilidad.

Tabla 16 Pruebas de Verosimilitud, Wald y Score.

Concordance= 0.8	(se = 0.088)
Rsquare= 0.166	(max possible= 0.964)
Likelihood ratio test= 18.87	on 3 df, p=0.0002906
Wald test = 10.51	on 3 df, p=0.01471
Score (logrank) test = 15.47	on 3 df, p=0.001456

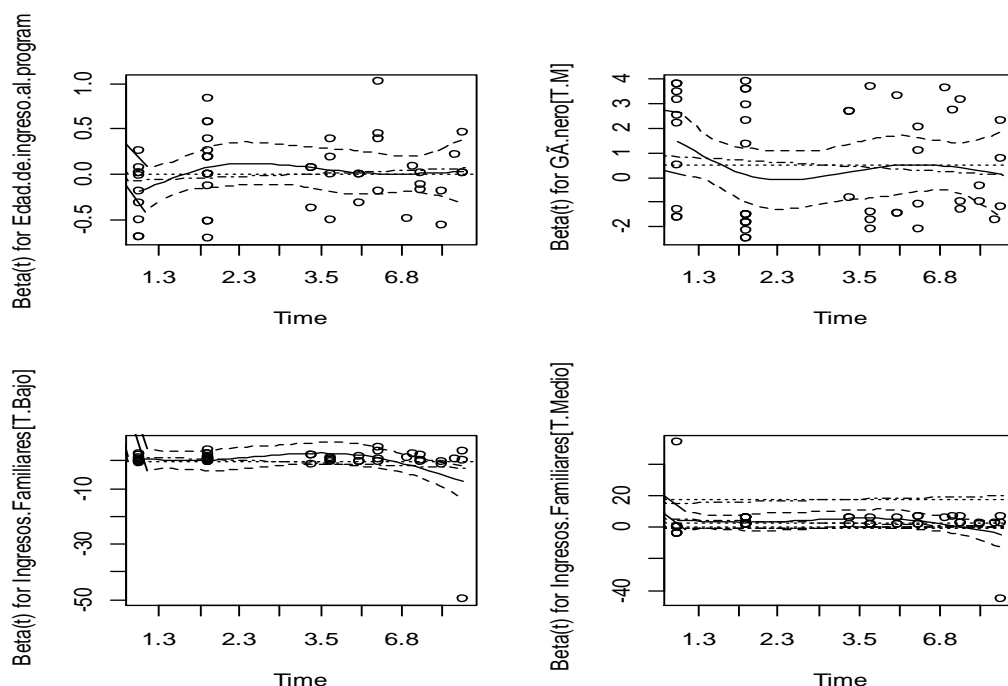
Fuente: Autor

4.3.3. Evaluación del Modelo

Con relación a los supuestos de riesgos proporcionales se viola de manera global, con una significancia del 5%, excepto para la covariable Ingresos familiares, esto se ve trazado en los residuos de Schoenfeld en la gráfica 18, al denotarse una línea suave por cada covariable.

	rho	chisq	p
Número.de.materias.repetidas.	0.419	1.09e+01	0.000942
Puntaje.estandarizado.ICFES	0.283	4.67e+00	0.030649
Ingresos.Familiares[T.Bajo]	0.610	9.58e-08	0.999753
GLOBAL		1.15e+01	0.021744

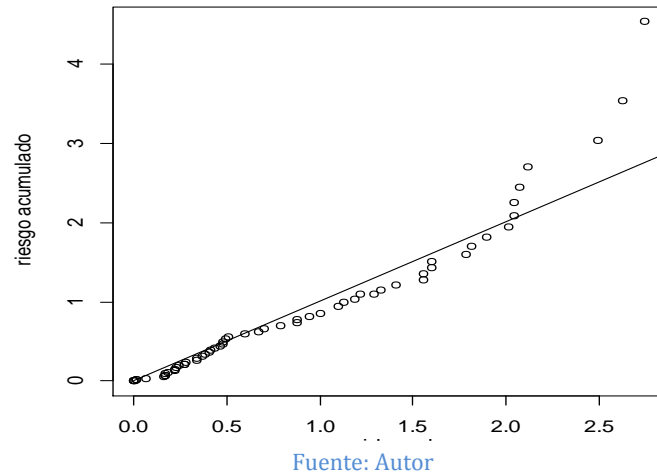
Gráfica 18 Residuos de Schoenfeld



Fuente: Autor

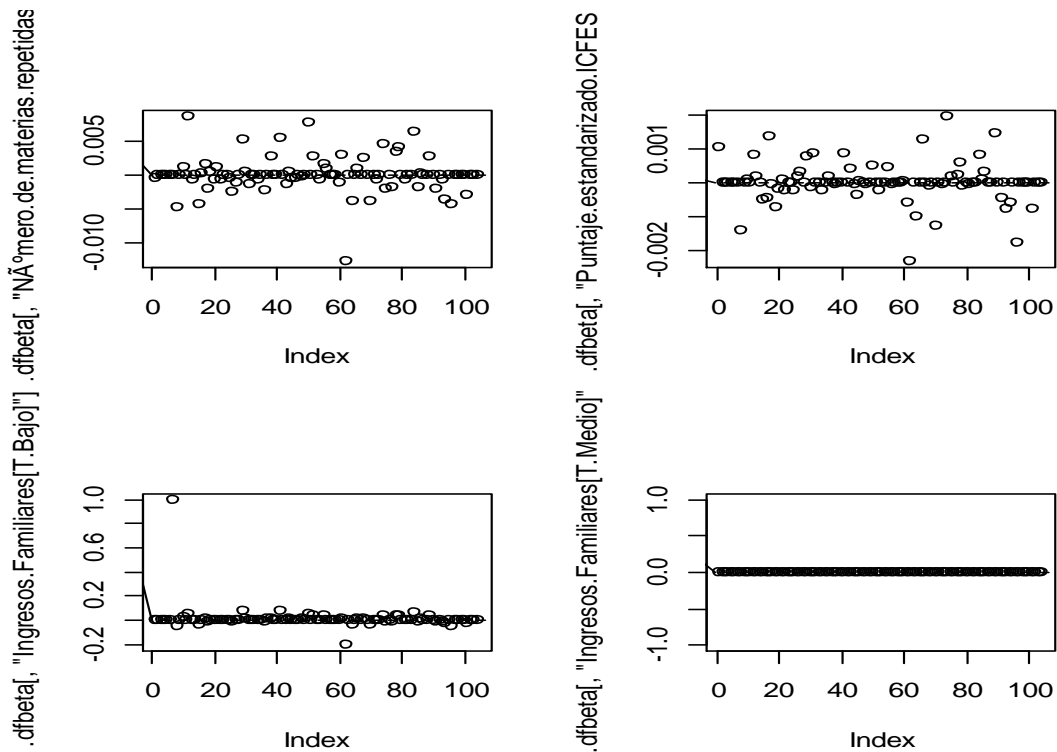
Residuos de Cox-Snell: De manera general se comprueba la hipótesis de riesgos proporcionales, ya que se ajustan a la línea recta, y en lo que tiene que ver con los residuos de la derecha se presentan por la censura de los datos presentados para este programa.

Gráfica 19 Residuos de Cox-Snell para el modelo de Graduación en AEA



Residuos tipo dfbeta: En este caso ninguna de las observaciones ejerce un cambio mayor a aproximadamente 0.005% del error estándar como se observa en la gráfica 20. Si se comparan con el valor coeficiente de regresión estimado y su error estándar podemos concluir que ninguna de las observaciones es influyente individualmente.

Gráfica 20_Residuos tipo dfbeta



Fuente: Autor

Como conclusión principal del programa de Administración de Empresas Agropecuarias, se tiene que la deserción es del 46%, cifra que según el reporte del MEN (2014) supera la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2%, sin embargo en cuanto a la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines es del 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 46, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría (46,45%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Pero también es considerable la proporción de aquellos que lo hacen por causas no académicas predominado aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan de igual forma los hombres y mujeres, en el momento de presentar el ICFES no trabajaban, que además reportan no tener vivienda propia, cuya madre alcanza el nivel educativo de primaria, y sus ingresos familiares son de nivel bajo. El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, es decir del segundo al quinto semestre. Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Los factores que aumentan el riesgo de desertar son los de ser hombre y el de tener ingresos en nivel medio, hecho que es consistente con lo planteado por (Giovagnoli, 2002) quien lo atribuye a factores culturales o bien a que el hombre

es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa y al registrar ingresos familiares en nivel bajo, hecho particular que refleja que este tipo de estudiantes son mucho más comprometidos al hacer un esfuerzo mayor de avanzar dentro del Programa, posiblemente abarcando las oportunidades que ofrece la Universidad. Con base en lo anterior, la Universidad y el Programa en particular deberían diseñar acciones preventivas de acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración de Empresas agropecuarias es de 20 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas y tener un puntaje alto en el ICFES disminuyen la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. Por el contrario registrar ingresos familiares en nivel bajo o medio, aumentan esta probabilidad de graduarse.

5. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN INDUSTRIAL

En este capítulo se describe el proceso que se empleó para elaborar cada modelo de sobrevivencia, incluyendo un análisis de datos (de forma univariada y bivariada), posteriormente se referencian los supuestos del modelo y se muestra el modelo para la deserción y graduación, en donde para cada uno, se emplean el estimador de la función de sobrevivencia Kaplan-Meier, función de riesgo acumulado, luego se estima el modelo de regresión de Cox y por último se validan los supuestos de este.

5.1. DESCRIPCIÓN DE LOS DATOS

Esta sección muestra una descripción de forma univariada cada una de las variables de interés. Algunas de ellas se refieren a aspectos personales, familiares, académicos e institucionales, en donde se puede ver resumidos algunos estadísticos como las media, la mediana, mínimo y máximo. En segundo lugar, se hace el análisis bivariado a través de tablas de contingencia y pruebas de correlación de Pearson. La tabla 17 muestra el resumen numérico y descriptivo de las variables objeto de estudio y otras de interés.

Tabla 17 Resumen descriptivo de las variables de estudio del programa de AI

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Género	F= Mujer M= Hombre	110 (61,8 %) 68 (38,2 %)
Nivel educativo de la madre	Básica Primaria = BP Básica Secundaria = BS Técnico = TEC Universitario = UN Sin Clasificar = SC	64 (35,955 %) 64 (35,955 %) 10 (5,617 %) 13 (7,30 1 %) 27 (15,16 %)
Número de Hermanos	1 = Uno 2= Dos 3 = Tres 4 = Cuatro 5 = Cinco Sin Clasificar = SC	16 (8,98 %) 78 (43,82 %) 37 (20,78 %) 12 (6,74 %) 5 (2,80 %) 30 (16,85 %) Media = 2.405 Min – Max = 1-5 Mediana = 2

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Edad de Ingreso al programa	En años cumplidos	Media = 19 Min – Max = 15 - 28 Mediana = 19
Ingreso Familiar al Presentar el Examen de Estado	Bajo Medio Alto Sin Clasificar= SC	123 (69,10%) 21 (11,79%) 6 (3,37 %) 28 (15,73 %)
Trabajaba al presentar el examen de estado	Si No Sin clasificar = SC	3 (1,69 %) 149 (83,71 %) 26 (14,06 %)
Vivienda Propia	Si = Si No = No Sin Clasificar = SC	111 (62,36 %) 52 (29,21 %) 15 (8,43 %)
Puntaje en el Examen de Estado	De 1 – 100 puntos	Media = 69,47 Min – Max = 10 - 99 Mediana = 72
Estado del estudiante	Matriculado Desertor Graduado	16 (8,99 %) 89 (50 %) 73 (41,01 %)
Número de Materias Perdidas		Media = 7,5 Min – Max = 0 - 32 Mediana = 6
Número de Materias Repetidas		Media = 6,79 Min – Max = 1 - 24 Mediana = 6
Tiempo Hasta la Deserción	Semestres Académicos (16 semanas)	Media = 8.51 Min – Max = 1 - 13 Mediana = 12
Estado del estudiante Dicotomizado	Desertor No Desertor	89 (50%) 89 (50 %)
Tipo de deserción	Académica No Académica	36 (40,44%) 53 (59,55%)
Motivo de deserción Académica “Perdió Cupo/ Artículo 80”	Literal B	9 (25%)
	Literal C	14 (38,88 %)
	Literal D	9 (25%)
	Literal E	4 (11,11 %)
Tipo de Deserción	Precoz	29 (32,58 %)
	Temprana	48(53,93 %)
	Tardía	12 (14,48 %)

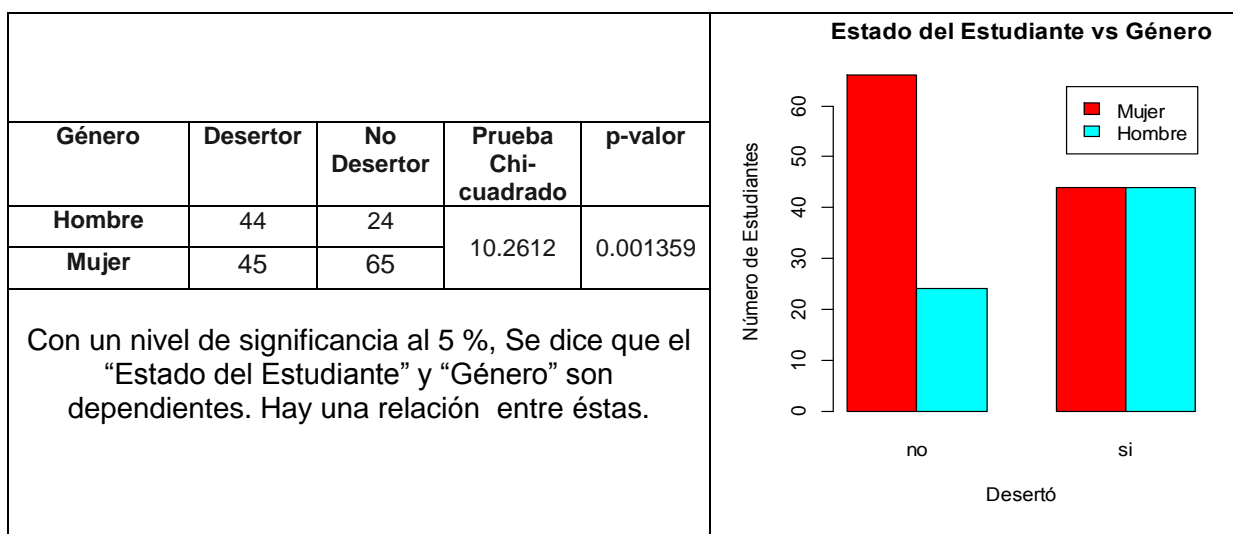
Fuente: Autor

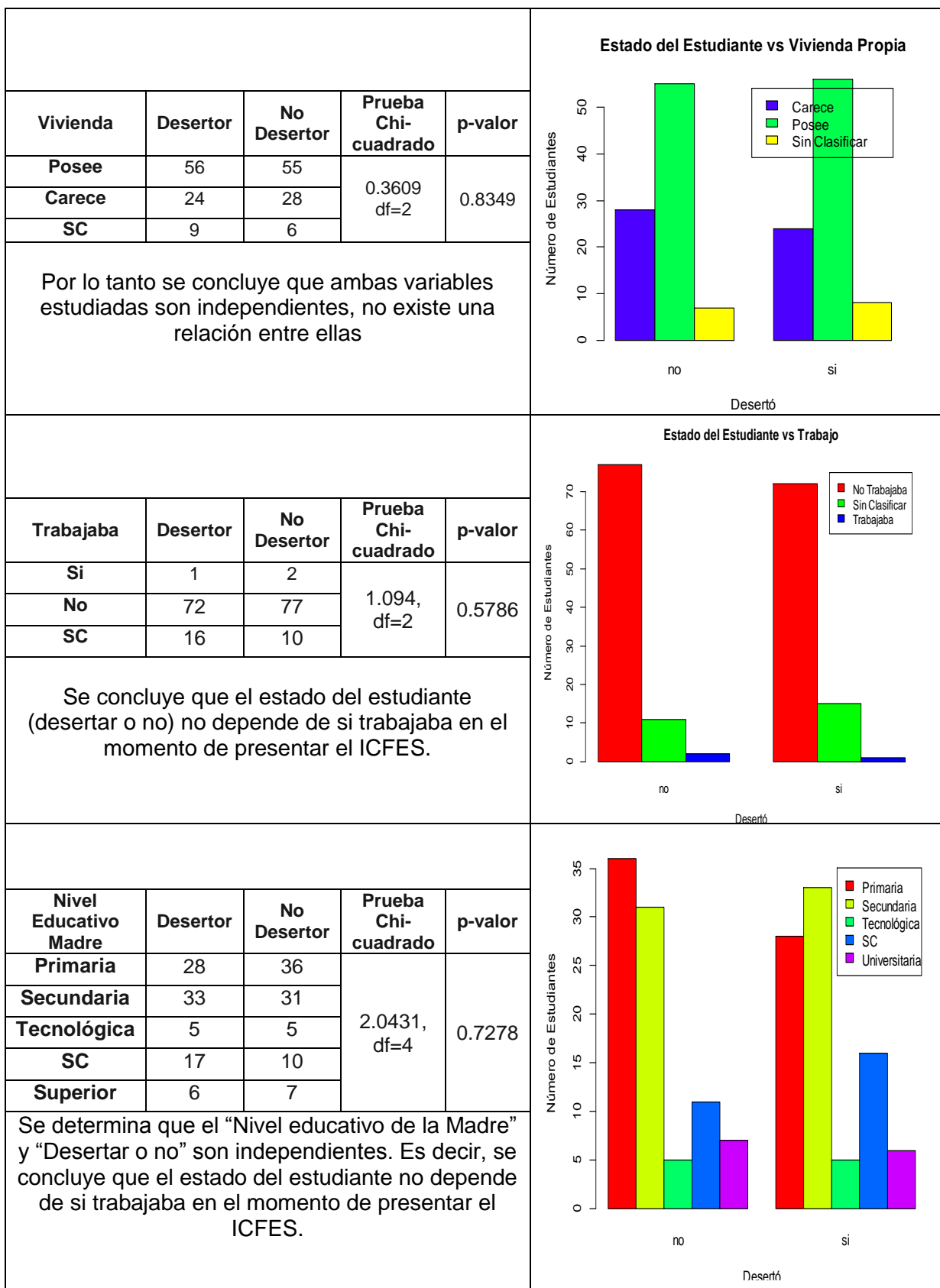
Con los resultados anteriores se puede observar que de los estudiantes seleccionados en la muestra, la mayoría son mujeres (61,8 %). La edad media al momento de ingresar al programa fue de 19 años. El 62,32% tenían vivienda propia y el 83,71% no trabajaban al momento de presentar el examen de Estado.

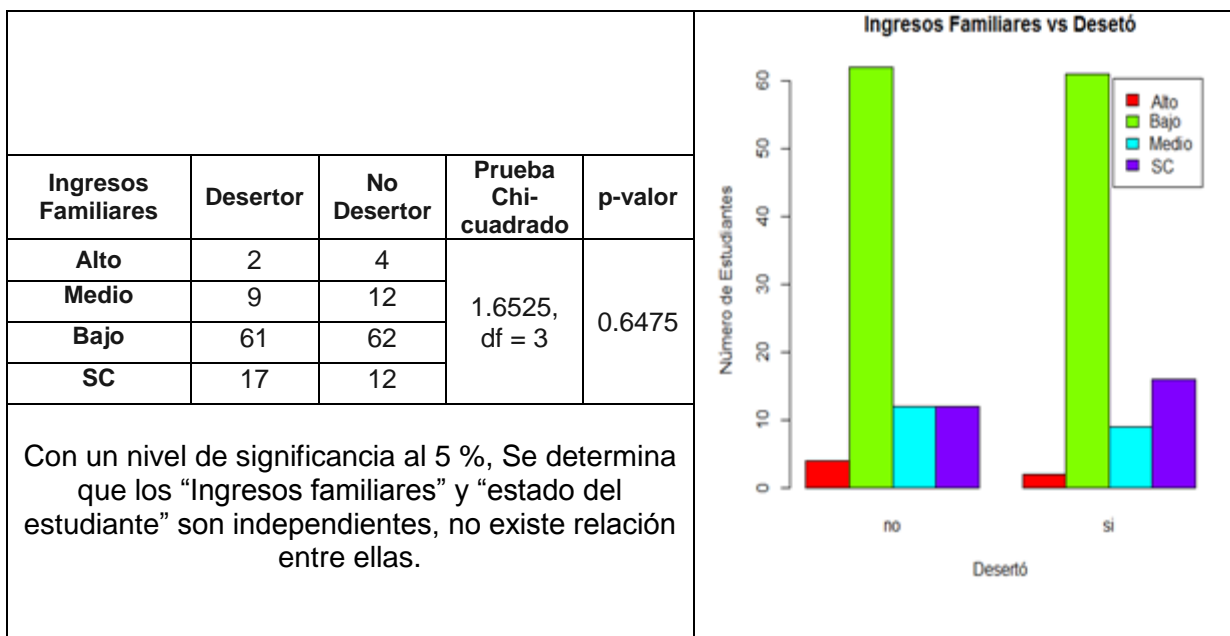
El puntaje promedio al momento de presentar el examen de estado fue de 69 en una escala de 0 a 100. El 77,5 % tenía como ingreso familiar de nivel bajo. Para este conjunto de datos, se tiene que al primer semestre de 2015, ha desertado el 50% de los estudiantes de las cohortes entre el primer semestre de 2004 y el segundo semestre de 2009, donde el 53,93% presentan deserción temprana. En estos mismos tiempos, se ha graduado el 41 % de los estudiantes. Además el 64 % de los estudiantes tenían madres con nivel educativo de básica primaria. El 43,82 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

A continuación en la tabla 18 se presenta una descripción bivariada del estado del estudiante, mediante tablas de contingencia como se evidencia en la tabla 18:

Tabla 18 Contrastes de algunas variables categóricas del programa de AI







Fuente: Autor

Ahora se realiza el test de correlación de Pearson entre tiempo hasta deserción y variables cuantitativas continuas: para comprobar si el tiempo hasta la deserción de los estudiantes está correlacionada o no, como se resume en la tabla 19:

Tabla 19 Test de correlación de Pearson

Variable	Tiempo hasta Desertar		
	Coefficiente de Correlación	Estadístico t	p – valor
Edad de Ingreso al Programa	-0.05714135	-0.578	0.5645
Número de Materias Perdidas	0.2586593	2.7044	0.00802
Número de Materias Repetidas	0.2607583	2.7279	0.007506
Puntaje Icfes	0.04076746	-0.4121	0.6812
Número de Hermanos	-0.01641232	-0.1658	0.8687

Fuente: Autor

Se observa que las variables correlacionas con el tiempo hasta Desertar son el número de materias pérdidas y número de materias repetidas, con una correlación leve, a medida que aumenta el tiempo hasta desertar también aumenta el número de materias perdidas y el número de materias repetidas.

5.2. MODELO DE SOBREVIDA PARA DESERCIÓN EN AI

Como el propósito del estudio es identificar los factores que están relacionados al tiempo de la deserción, en primer lugar, se presentará la aplicación de la metodología de Kaplan Meier para estimar la función de sobrevivida, con la cual se puede analizar la evolución de la probabilidad de deserción con su respectivo intervalo de confianza. Posteriormente, se construirá el modelo de regresión de Cox para estimar el efecto de las variables de estudio sobre los tiempos de sobrevivencia a la deserción. Recordemos las características del estudio:

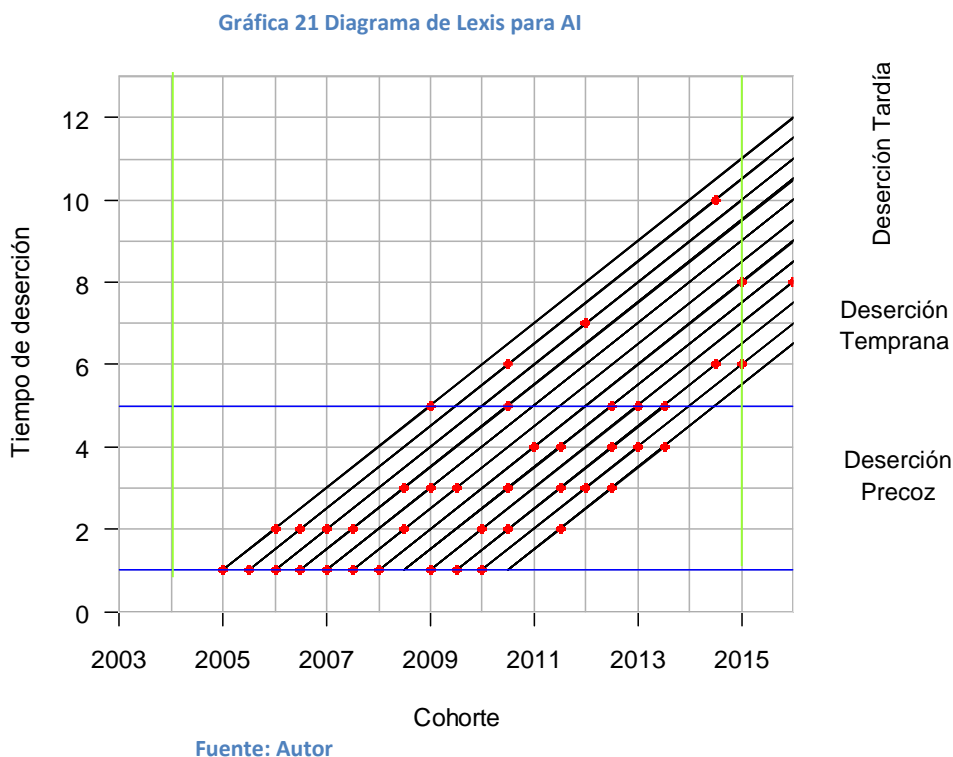
Tabla 20 Consideraciones para el Modelo de Sobrevida para AI

La unidad sobre la cual se registra el evento		Estudiantes seleccionados en la muestra, matriculados en cada programa de administración industrial en primer semestre.
Evento de interés		Desertar del Programa
Variable respuesta		Tiempo hasta la Deserción: Cuya escala de medición es de razón de tipo discreto, ya que se mide en número de semestres cursados hasta que presenta el fenómeno de estudio.
Tiempo	Inicial del Estudio	I Semestre de 2004
	Origen del Evento	Cohorte del Estudiante (primera matricula del estudiante por programa)
	Final del Estudio	I Semestre de 2015
La escala de medida del tiempo hasta el evento		Semestres Calendario(16 Semanas)
Censura	Tipo	Tipo I (a derecha)
	Tiempo	13 semestres por cohorte

Fuente: Autor

En seguida se presenta la representación de los estudiantes del estudio mediante el diagrama de Lexis, (ver sintaxis en R en el anexo E), este evidencia que los estudiantes bajo estudio no tienen el mismo tiempo de origen, se señala con los puntos rojos aquellos que presentaron el evento y el tiempo en que sucedió (expresado en semestres).

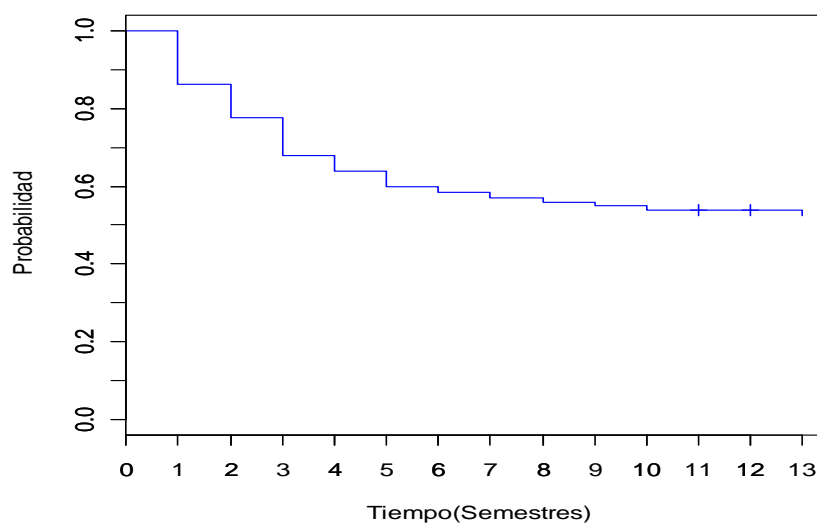
Nótese que la mayoría de estudiantes desertan precozmente, seguido de la deserción temprana, unos pocos se registran con deserción tardía.



5.2.1. Función de Sobrevida

Esta curva muestra en la gráfica 22 y la tabla 21 que la probabilidad de continuar con el programa va disminuyendo a medida que el estudiante avanza durante el programa. Nótese que es más rápido el decrecimiento en los primeros seis semestres y tiende a estabilizarse gradualmente a partir del 6 semestre donde está alrededor de 58,5%. Por ejemplo la probabilidad de continuar con el programa hasta el primer semestre es del 86,4% y baja a 77,6% en el segundo semestre, luego en el tercer semestre esta posibilidad es del 68%, llegando así con un 52,5% al semestre 13.

Gráfica 22 Función de sobrevivencia para AI



Fuente: Autor

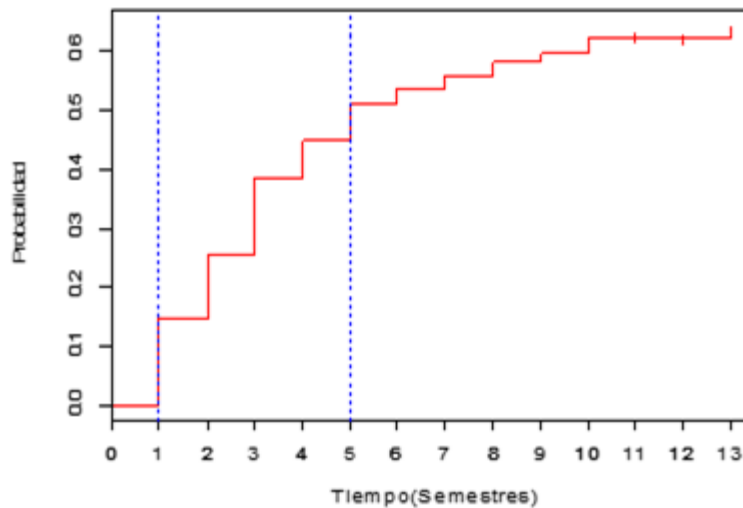
Tabla 21 Estimador de KM en AI

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	147	20	0.864	0.0283	0.810	0.921
2	127	13	0.776	0.0344	0.711	0.846
3	114	14	0.680	0.0385	0.609	0.760
4	100	6	0.639	0.0396	0.566	0.722
5	94	6	0.599	0.0404	0.524	0.683
6	88	2	0.585	0.0406	0.511	0.670
7	86	2	0.571	0.0408	0.497	0.657
8	84	2	0.558	0.0410	0.483	0.644
9	82	1	0.551	0.0410	0.476	0.638
10	81	2	0.537	0.0411	0.463	0.624
13	45	1	0.525	0.0419	0.449	0.614

Fuente: Autor

La función de riesgo acumulado, permite tener información del comportamiento del riesgo a lo largo del tiempo, a continuación se presenta la correspondiente al programa de AI, para efectos de interpretación se indica en la gráfica el semestre 1 (deserción precoz) y semestre 5 (deserción temprana).

Gráfica 23 Función de Riesgo acumulada de AI

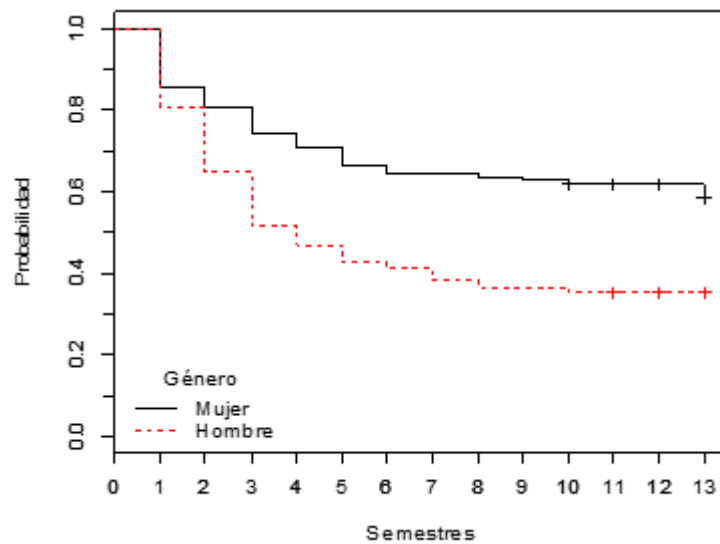


Fuente: Autor

Según la función de riesgo acumulada se puede decir que la probabilidad que un estudiante deserte precozmente en el primer semestre es de 15% siendo la más alta, de igual manera se observa que la función de riesgo presenta los mayores cambios en la deserción temprana, por ejemplo, en este intervalo de tiempo (1-5 Semestre) se tiene un punto crítico de que habiendo llegado a segundo deserte en el tercer semestre es del 25,3%. Además la posibilidad de que llegado al tercero deserte en el cuarto es alrededor de 38,5%.

Como el fenómeno de la deserción no se está estudiando de manera aislada, se hace necesario un análisis de la función de sobrevivencia separado por poblaciones de estudio, lo cual se presenta a continuación. En primer lugar, veremos la función de sobrevivencia por género, tanto las funciones como el resumen numérico se muestra:

Gráfica 24 Función de Sobrevida por Genero para AI



Fuente: Autor

Esta curva de sobrevivencia evidencia que tienen mayor posibilidad de desertar los hombres, esta tendencia está marcada a partir del segundo semestre, y llegando al último semestre con una probabilidad de sobrevivir cercana del 34,6%, las mujeres denotan mayor posibilidad de culminar con sus estudios llegando así en el semestre 13 con cerca del 62.1 %, un 28% de probabilidad más que los hombres.

Tabla 22 Estimador de KM por género

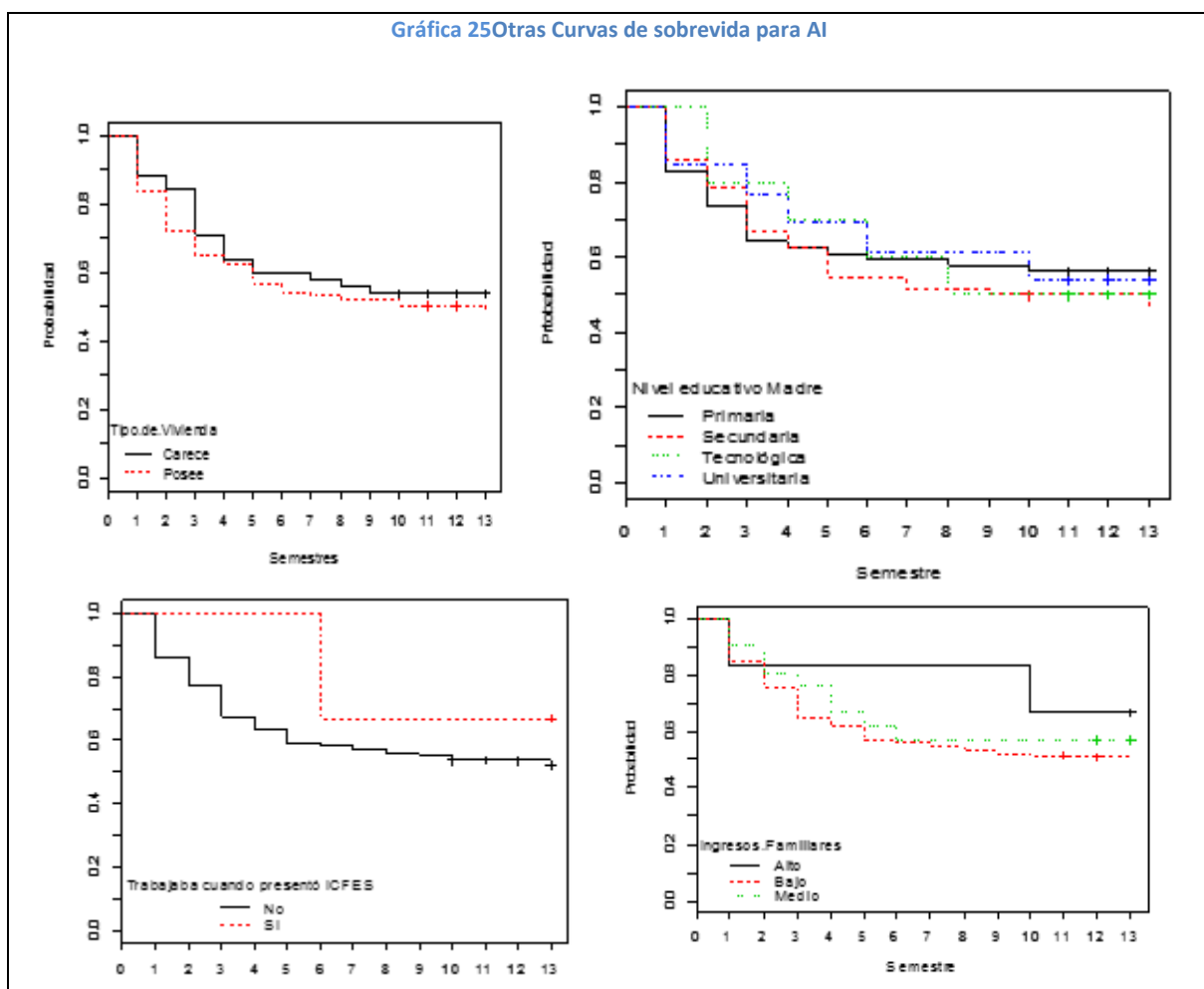
Género=Femenino					95% CI		Género=Masculino					95% CI	
time	n.risk	n.event	Survival	std.err	lower	Upper	time	n.risk	n.event	survival	std.err	Lower	Upper
1	95	10	0.895	0.0315	0.835	0.959	1	52	10	0.808	0.0547	0.707	0.922
2	85	5	0.842	0.0374	0.772	0.919	2	42	8	0.654	0.0660	0.537	0.797
3	80	7	0.768	0.0433	0.688	0.858	3	34	7	0.519	0.0693	0.400	0.674
4	73	4	0.726	0.0457	0.642	0.822	4	27	2	0.481	0.0693	0.362	0.638
5	69	3	0.695	0.0472	0.608	0.794	5	25	3	0.423	0.0685	0.308	0.581
6	66	2	0.674	0.0481	0.586	0.775	7	22	2	0.385	0.0675	0.273	0.542
8	64	1	0.663	0.0485	0.575	0.765	8	20	1	0.365	0.0668	0.255	0.523
9	63	1	0.653	0.0489	0.564	0.756	10	19	1	0.346	0.0660	0.238	0.503
10	62	1	0.642	0.0492	0.553	0.746							
13	31	1	0.621	0.0518	0.528	0.732							

Fuente: Autor

Lo anterior se corrobora al obtener un valor de Chisq = 12.2 con 1 grado de libertad y un p valor de 0.00469, estadística que permite afirmar que se rechaza la hipótesis nula de igualdad de las curvas de sobrevivida afirmando que la sobrevivencia (a la deserción) de hombres y mujeres no tiene el mismo comportamiento. Al efectuar las pruebas de LonngRank para nivel educativo de la madre, ingreso familiar y tenencia de vivienda, se encontró que no se rechaza la hipótesis nula de igualdad de las funciones tal como se evidencia en la gráfica 25, para cada una de estas se resume así:

Variable	Chisq	Grados de Libertad	P valor
Nivel.educativo.madre	0.7	3	0.866
Tipo.de.Vivienda	0.1	1	0.795
Trabajaba.presento.ICFES	0.3	1	0.554
Ingresos.Familiares,	0.7	2	0.691

Gráfica 25 Otras Curvas de sobrevivida para AI



Fuente: Autor

5.2.2. Factores relacionados con el riesgo de Desertar

En el apartado anterior se hizo un análisis de la situación de sobrevivencia de los estudiantes del programa de AI, en esta sección se pretende llegar a explicar los factores relacionados con el riesgo de deserción en los estudiantes mediante el modelo de regresión de Cox. Inicialmente se presenta el modelo completo con todas las variables, luego para la obtención del mejor conjunto de variables explicativas del modelo se tiene en cuenta el Criterio de Información de Akaike (AIC) con el método de selección hacia adelante. Posteriormente se calcula la curva de Sobrevivencia. Las variables a considerar son el género, la edad de ingreso al programa, nivel de ingresos familiares, nivel educativo de la madre, puntaje estandarizado del ICFES, estado de empleo cuando presentó el ICFES, tenencia vivienda.

Para la selección del modelo primero se muestra el modelo de regresión de Cox inicial con todas las covariables:

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-0.071436	0.931056	0.066467	-1.075	0.28248
Trabajaba.cuando.presento.ICFES[T.Si]	-0.979724	0.375415	1.030407	-0.951	0.34170
Género[T.M]	0.886814	2.427383	0.258074	3.436	0.00059 ***
Ingresos.Familiares[T.Bajo]	0.919701	2.508539	1.023429	0.899	0.36884
Ingresos.Familiares[T.Medio]	0.569551	1.767474	0.977812	0.582	0.56025
Nivel.educativo.madre[T.Secundaria]	0.094308	1.098898	0.335323	0.281	0.77852
Nivel.educativo.madre[T.Tecnológica]	0.310916	1.364675	0.627666	0.495	0.62035
Nivel.educativo.madre[T.Universitaria]	0.444627	1.559909	0.792821	0.561	0.57492
No..De.hermanos	-0.166352	0.846748	0.179591	-0.926	0.35430
Puntaje.estandarizado.ICFES	-0.008854	0.991186	0.006643	-1.333	0.18264
Tipo.de.Vivienda[T.Posee]	0.018090	1.018255	0.297034	0.061	0.95144

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Seguidamente se procede a seleccionar le modelo más parsimonioso, utilizando el método hacia adelante y usando el criterio AIC (en el Anexo D se muestra en detalle esta selección), encontrándose que el modelo óptimo queda determinado por el género y la edad de ingreso al programa, cómo se evidencia a continuación:

n= 147, number of events= 69

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-0.05195	0.94938	0.06099	-0.852	0.39435
Género[T.M]	0.83617	2.30752	0.24199	3.455	0.00055 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Edad.de.ingreso.al.programa	0.9494	1.0533	0.8424	1.070
Género[T.M]	2.3075	0.4334	1.4360	3.708

En cuanto a las características observadas en los estudiantes se tiene tan sólo el género es significativa a la hora de explicar el riesgo de deserción.

En la edad de ingreso al programa el riesgo relativo es de 0.945, con un intervalo de confianza al 95% (0.8424, 1.070). Como la estimación puntual es inferior a 1, indica que por cada año de edad de ingreso al programa disminuye el riesgo de desertar, dejando la otra variable constante, concretamente si se comparan dos estudiantes, aquel estudiante mayor en un año de edad tiene 1.0533 veces más probabilidad de desertar.

Además se tiene que los Hombres tienen un riesgo de desertar de 2.3075 veces mayor que de las mujeres, hecho que es consistente con otros estudios, donde afirman que las mujeres poseen mayor responsabilidad y perseverancia (Giovangnoli, 2002).

Es decir que el factor que aumenta el riesgo de desertar en AI es ser hombre, y este disminuye a una mayor edad de ingreso al programa.

Se observa en la tabla 23 que en cuanto a la bondad de ajuste del modelo, para las tres pruebas se aprecia un p-valor significativamente pequeño, lo cual es evidencia de que estas covariables explicativas al menos una tiene un coeficiente distinto de cero en el modelo de riesgos proporcionales, además que las variables

seleccionadas para el estudio tan sólo explican el 16.2% de la variabilidad en el tiempo de deserción.

Tabla 23. Pruebas de Verosimilitud, Wald y Score.

Concordance= 0.698 (se = 0.052)
Rsquare= 0.162 (max possible= 0.94)
Likelihood ratio test= 19.26 or 2 df, p=0.001721
Wald test = 15.21 or 2 df, p=0.009504
Score (logrank) test = 19.14 or 2 df, p=0.001812

Fuente: Autor

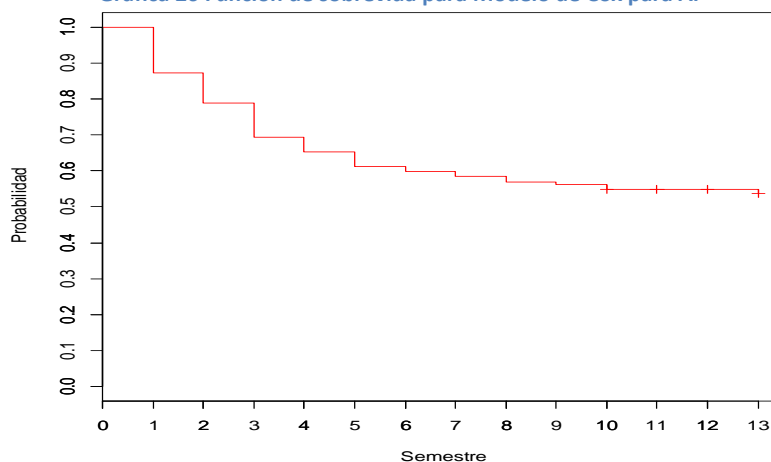
Se presenta la función de supervivencia para el modelo estimado en la tabla 24 y la curva de supervivencia para el modelo de Cox Ajustado de AI en la gráfica 26. Se tiene que la probabilidad de continuar en el Programa al finalizar el primer semestre es del 86,6%, en el segundo baja al 79% llegando al semestre trece con un 53.3% de probabilidad de seguir en este.

Tabla 24 Función de Supervivencia modelo de Cox en AI

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	147	20	0.876	0.0268	0.825	0.930
2	127	13	0.790	0.0337	0.727	0.859
3	114	14	0.694	0.0388	0.622	0.775
4	100	6	0.652	0.0403	0.578	0.736
5	94	6	0.610	0.0415	0.534	0.697
6	88	2	0.596	0.0418	0.519	0.684
7	86	2	0.582	0.0420	0.505	0.670
8	84	2	0.567	0.0422	0.490	0.657
9	82	1	0.560	0.0423	0.483	0.650
10	81	2	0.546	0.0425	0.469	0.636
13	45	1	0.533	0.0434	0.455	0.626

Fuente: Autor

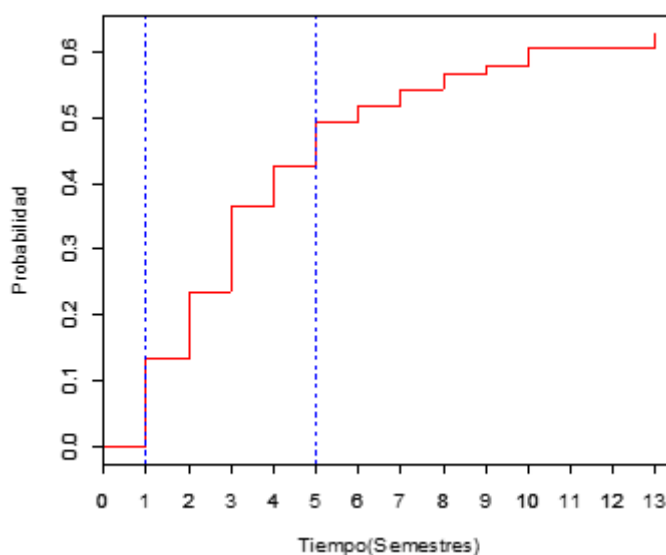
Gráfica 26 Función de supervivencia para Modelo de Cox para AI



Fuente: Autor

Según la observación de la gráfica 27 se tiene que el riesgo de que un estudiante deserte justo al terminar el primer semestre es del 13,23 %, en donde hay varios puntos críticos por ejemplo, de que la probabilidad de que habiendo llegado a primero deserte en el segundo semestre es del 23,5% y que llegando al segundo deserte en el tercer semestre es del 30% %además se tiene un punto crítico en segundo semestre donde la probabilidad de que habiendo llegado a segundo deserte en el tercer semestre es del 36,5 %, Donde el ajuste que se obtiene de las probabilidades corresponde a la inclusión en su cálculo de la edad de ingreso y el género.

Gráfica 27 Función de riesgo en modelo de deserción en AI



Fuente: Autor

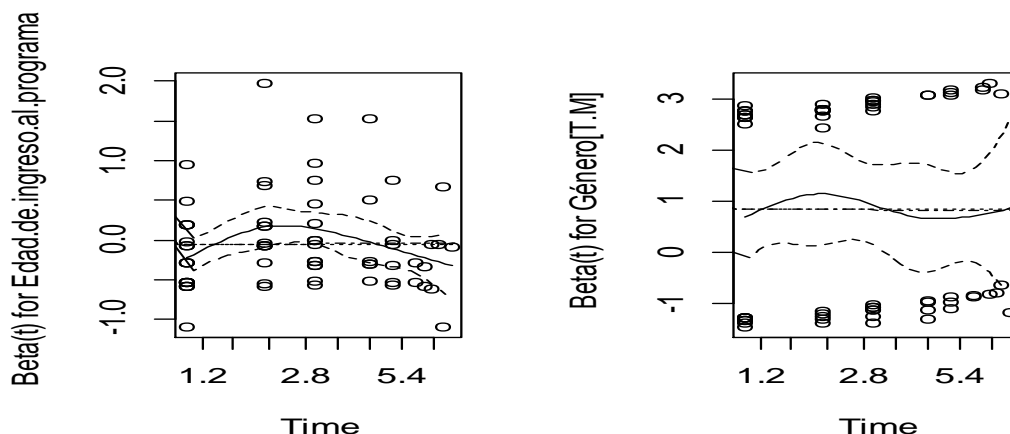
5.2.3. Evaluación del modelo de riesgos proporcionales

En seguida se verifica el supuesto de riesgos proporcionales del ajuste del modelo de regresión, como sigue:

	rho	chisq	p
Edad.de.ingreso.al.programa	0.01210	0.013150	0.909
Género[T.M]	-0.00245	0.000415	0.984
GLOBAL	NA	0.013385	0.993

Por lo que, no existen evidencias significativas al 5% de que se viole el supuesto de riesgos proporcionales para ninguna de las covariables ni globalmente. Además gráficamente los residuos escalonados de Schoenfeld, trazan una curva suave para las covariable, como se refleja en seguida:

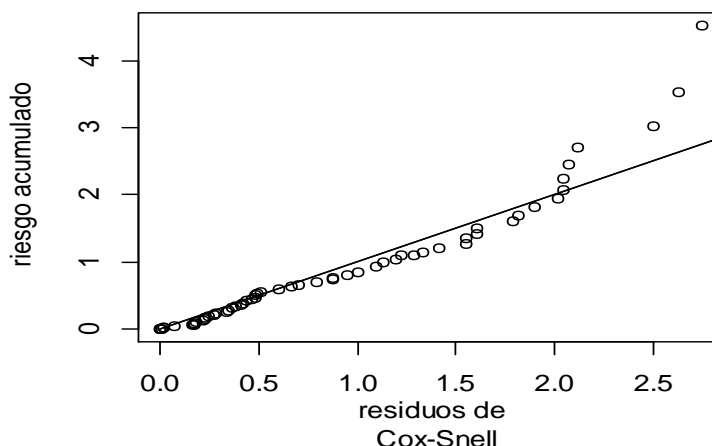
Gráfica 28 Residuos escalados de Schoenfeld



Fuente: Autor

Comprobación de la hipótesis global del modelo mediante los residuos de Cox-Snell: Si el modelo es correcto y la estimación de los β son cercanas a los valores reales, entonces este estimador debería describir aproximadamente una recta de pendiente igual a la unidad. La gráfica 29 Indica que, en general, este modelo se ajusta bien a los datos, y las cola a derecha según los datos se presentan por censuras presentadas en el programa.

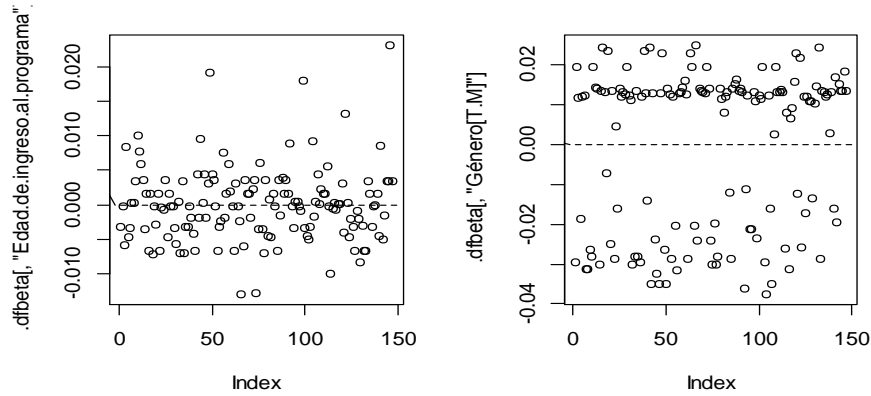
Gráfica 29 Residuos de Cox-Snell



Fuente: Autor

Residuos dfbeta: Con estos, se determina la influencia de cada observación en el modelo. De manera que, para cada covariable, se ha representado la observación (en orden de tiempo de falla registrado) por el cambio de escala aproximada del coeficiente, es decir, dividiendo por el error estándar, después de la eliminación de la observación del modelo.

Gráfica 30 Residuos Df Beta



Fuente: Autor

Se observa que los residuos no presentan una mayor variación al 2%, es decir que ningún individuo contribuye a que los valores de los coeficientes se alteren.

5.2.4. Pronóstico a partir del modelo

En la sección 2.4.5 se mencionó que la función de riesgo inicial $h_0(t)$ depende del tiempo, mientras que $\exp(X\beta^T)$ depende únicamente de las covariables. Sin embargo, en el modelo de riesgos proporcionales de Cox se debe tener en cuenta que los riesgos para dos conjuntos diferentes de valores de los covariables conservan la misma proporción a lo largo del tiempo. Con base en los resultados obtenidos en el modelo de regresión de deserción en AI, este corresponde a:

$$h_i(t, X, \beta) = h_0(t) \exp(-0.05195(\text{Edadingreso}) + (0.83617(\text{Género}[\text{T.M}])) \quad (5.25)$$

El exponente del modelo de Cox, que particularmente para el estudio corresponde a la ecuación (5.25) se denomina *índice pronóstico* (P_I), (Domn  ch, 1992). De manera que la tasa de riesgo se puede expresar a trav  s de este indicador en lugar utilizar los valores del estudiante para cada variable. En la pr  ctica es mejor usar el *  ndice de pron  stico centrado* (P_{IC}), que se obtiene centrando las variables predictoras:

$$P_{IC} = \beta_1(X_1 - \bar{X}_1) + \dots + \beta_p(X_p - \bar{X}_p)$$

Que para el programa de AI, corresponde a:

$$P_{IC} = (-0.05195((X_1 - 19) + (0.83617G  nero[T.M]))$$

Un   ndice pron  stico centrado igual a cero de AI corresponde a una estudiante mujer con edad promedio de 19 a  os, y que denominaremos de ahora en adelante “estudiante base”. La diferencia entre los   ndices pron  sticos de dos estudiantes permite estimar su riesgo relativo (RR), veamos:

$$RR = \frac{h_i(t, X_i, \beta)}{h_j(t, X_j, \beta)} = \frac{h_0(t) \exp(P_{ICi \text{ individuo } i})}{h_0(t) \exp(P_{ICi \text{ individuo } j})} = \exp(P_{ICi \text{ individuo } i} - P_{ICi \text{ individuo } j})$$

Teniendo en cuenta lo anterior a continuaci  n se presenta el RR para cinco estudiantes, seleccionados aleatoriamente, y que se encuentran cursando el primer semestre del programa, es decir, estudiantes que ingresaron en el I semestre de 2015.

Tabla 25 Pron  stico del modelo de deserci  n para AI

N��	Estudiante		P_{IC}	RR respecto a un estudiante promedio
	G��nero	Edad.Ingr		
1	F	24	0.25975	1.2966
2	M	20	0.88812	2.4305
3	F	19	0	0
4	F	18	-0.05195	1.0533
5	F	17	-0.1039	1.1094

Lo anterior significa que respecto al estudiante base el riesgo de deserción para los estudiante 4 y 5 es menor en 1.0533 y 1.1094 veces, respectivamente. Mientras que para los estudiantes 1 y 2, el riesgo de desertar aumenta en 1.2966 y 2.4305, nótese que este riesgo es mayor en un hombre de 20 años de edad.

5.3. MODELO PARA LA GRADUACIÓN EN AI

La duración teórica de la carrera universitaria es de cinco años. Sin embargo, los alumnos que no desertan prolongan sus estudios por un mayor tiempo. Este apartado se concentra en una pregunta: ¿Qué factores “favorecen” a la finalización exitosa del programa de AI dado el tiempo que lleva el alumno estudiando?

Nótese que se sigue presentando censura tipo I a derecha, y el tiempo de censura se define a partir del *cuartil 3* del tiempo de permanencia de los estudiantes en el programa que corresponde a 14 semestres.

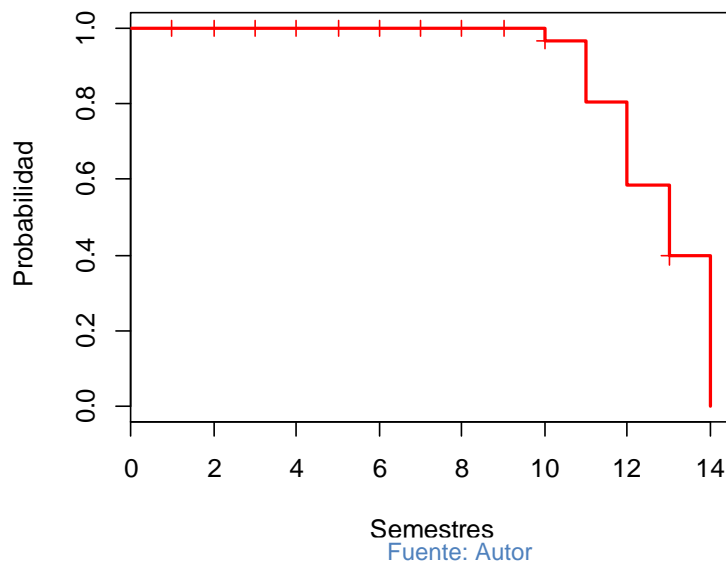
5.3.1. Función de sobrevida

En la tabla 26 y gráfica 31, se presenta la curva de sobrevida para la graduación en donde se observa que los eventos empiezan a darse desde el semestre 10 hasta el 14, la probabilidad de que un estudiante llegue a 10 semestre y no se haya graduado es de 96.8%, además llegado al semestre 11 y no haber presentado el evento es 80.3%, en el semestre 12 es de 58.3% la probabilidad de no haberse graduado y llegando al semestre 13 es de 39,6%.

Tabla 26 Estimador de KM para graduación

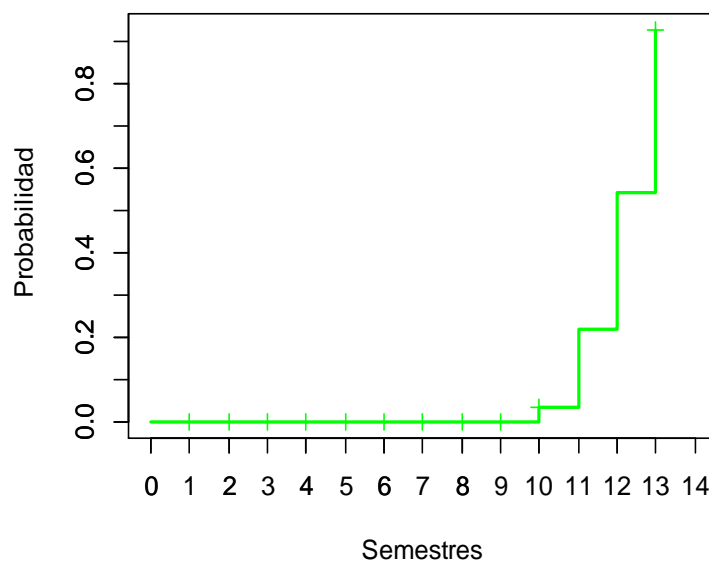
time	n.risk	n.event	survival	std.err	lower95% CI	upper 95%CI
10	93	3	0.968	0.0183	0.932	1.000
11	88	15	0.803	0.0417	0.725	0.889
12	73	20	0.583	0.0517	0.490	0.693
13	53	17	0.396	0.0513	0.307	0.510
14	35	35	0.000	NaN	NA	NA

Gráfica 31 Función de Sobrevida para graduación en AI



En la gráfica 32, se observa el riesgo de graduarse por ejemplo, en el décimo semestre es del 3.25%, por consiguiente de llegar al semestre número doce la probabilidad de graduarse es del 21.94% mientras que en el semestre trece la probabilidad de graduarse es del 92.6%.

Gráfica 32 Función de riesgo acumulada para la graduación



5.3.2. Factores asociados al aumento de graduación

En seguida se muestra el modelo de regresión de Cox para graduación con todas las covariables:

n= 147, number of events= 78

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-0.0276815	0.9726981	0.0685937	-0.404	0.687
Trabajaba.cuando.presento.ICFES[T.Si]	0.1556837	1.1684565	0.7929771	0.196	0.844
Género[T.M]	-0.1462224	0.8639655	0.3101434	-0.471	0.637
Ingresos.Familiares[T.Bajo]	0.2269108	1.2547180	0.9378077	0.242	0.809
Ingresos.Familiares[T.Medio]	0.1153580	1.1222751	0.8784841	0.131	0.896
Número.de.materias.repetidas.	-0.0538333	0.9475901	0.0725587	-0.742	0.458
Nivel.educativo.madre[T.Secundaria]	0.1603070	1.1738713	0.3617931	0.443	0.658
Nivel.educativo.madre[T.Tecnológica]	0.2737819	1.3149279	0.7584471	0.361	0.718
Nivel.educativo.madre[T.Universitaria]	1.0580535	2.8807581	0.8390674	1.261	0.207
No..De.hermanos	0.1221528	1.1299267	0.1826677	0.669	0.504
Puntaje.estandarizado.ICFES	-0.0003007	0.9996994	0.0069509	-0.043	0.965
Tipo.de.Vivienda[T.Posee]	-0.3809379	0.6832203	0.2854258	-1.335	0.182
X.Materias.perdidas.	-0.0216469	0.9785857	0.0611211	-0.354	0.723

La selección del modelo a través del criterio AIC y con el método de selección hacia adelante considera sólo una variable: número de asignaturas repetidas.

```

              coef exp(coef) se(coef)      z Pr(>|z|)
Número.de.materias.repetidas. -0.07206   0.93048  0.02187 -3.294 0.000988 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Número.de.materias.repetidas.   0.9305      1.075   0.8914   0.9712

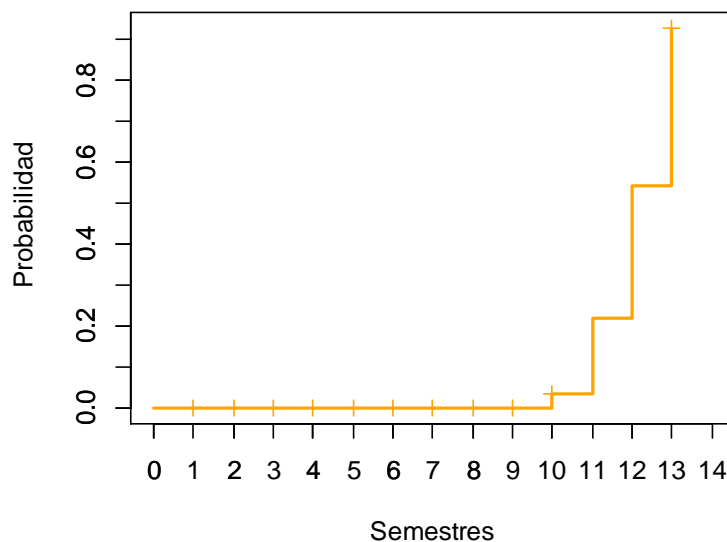
```

Los resultados indican que el riesgo relativo al número de materias repetidas es de 0.93, con un intervalo de confianza al 95%(0.89, 0.97). Como la estimación puntual es inferior a 1, señala que si se comparan dos estudiantes, para aquel estudiante que tenga una materia repetida de más disminuye en 1.075 la probabilidad de

graduarse. Nótese que la variable es significativa en el momento de explicar el aumento de los niveles de graduación.

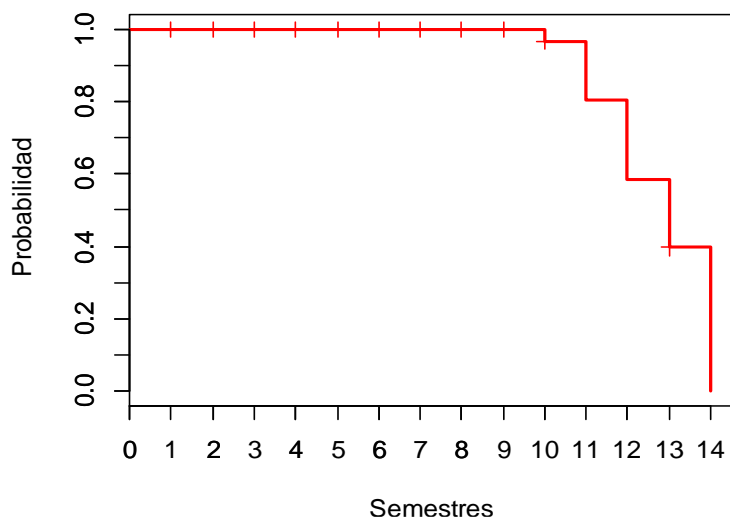
Se muestra a continuación la función de sobrevivencia para el modelo óptimo, la cual no presenta variaciones con función de supervivencia estimada de Kaplan-Meier.

Gráfica 33 Función de Sobrevivencia modelo de Graduación en AI



Fuente: Autor

Gráfica 34 Función de riesgo Modelo de Graduación en AI



Fuente: Autor

Para la bondad de ajuste del modelo se tiene que, el número de asignaturas repetidas es una variable explicativa con un coeficiente distinto de cero en el modelo de riesgos proporcionales bajo las tres pruebas, al momento de explicar la graduación en el programa de AI:

Tabla 27 Pruebas de Verosimilitud, Wald y Score.

Concordance= 0.768 (se = 0.076)			
Rsquare= 0.077 (max possible= 0.974)			
Likelihood ratio test=	11.82	on 1 df,	p=0.0005873
Wald test	= 10.85	on 1 df,	p=0.0009876
Score (logrank) test =	11.04	on 1 df,	p=0.0008913

Fuente: Autor

El modelo explica el 7,7% de la variabilidad en el tiempo de graduación.

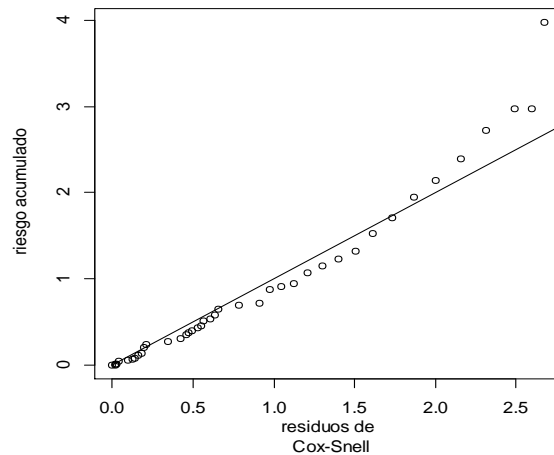
5.3.3. Evaluación del Modelo

Al efectuar la evaluación del modelo, se encuentra que el supuesto de riesgos proporcionales se viola para esta única covariable.

	rho	chisq	p
Número.de.materias.repetidas.	0.344	9.59	0.00196

Seguido de esto se calculan los Residuos de Cox-Snell, en donde se evidencia que el modelo se ajusta a los datos, luego la cola derecha se debe a las censuras presentadas por la información.

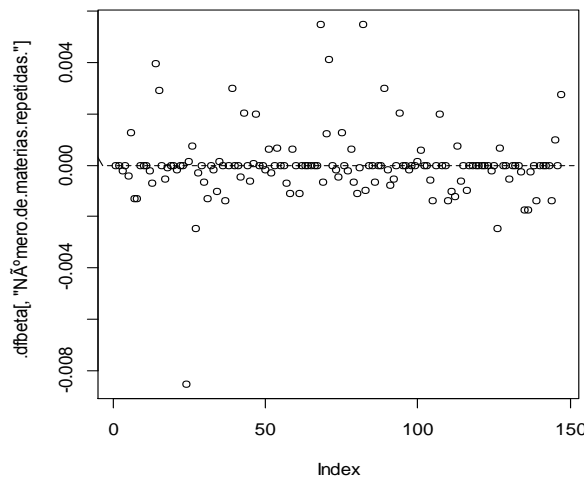
Gráfica 35 Residuos de Cox-Snell para el modelo de Graduación en AI



Fuente: Autor

Finalmente se grafican los residuos dfBetas: En donde de las observaciones no ejercen un cambio mayor a aproximadamente un 0.04% del error estándar. Si se comparan con el valor coeficiente de regresión estimado y su error estándar podemos concluir que ninguna de las observaciones es influyente individualmente.

Gráfica 36 Residuos Df Betas



Fuente: Autor

Como conclusión principal del programa de Administración Industrial se tiene que la deserción acumulada está cerca del 50%, cifra que supera según el reporte del MEN (2014) la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área

básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 50, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables. La mayoría (59.6%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Sin embargo, también es considerable la proporción (40.4%) de los estudiantes que salen por causas académicas predominando aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan mayoritariamente los que en el momento de presentar el ICFES no trabajaban y que además reportan tener vivienda propia, cuya madre alcanza el nivel educativo de secundaria, y sus ingresos familiares son de nivel medio.

El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, entre el segundo y quinto semestre. Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Al observar los factores que explican el hecho de desertar se encontró que aumenta el riesgo de desertar en ser hombre, hecho que va de la mano con lo que sucede a nivel Nacional. Una explicación a este hecho es la dada por Giovagnoli, 2002 en donde se le atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa. Con base en ese hecho la Universidad y el Programa en particular deberían diseñar acciones preventivas de

acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración Industrial es de 19 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas disminuye la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 13. Hecho que es preocupante ya que gasta tres semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. En este punto valdría la pena hacer un análisis de los posgrados que ofrece la Facultad y que sean afines al área de administración.

6. ANÁLISIS PARA EL PROGRAMA DE ADMINISTRACIÓN TURÍSTICA Y HOTELERA

En este capítulo se describe el proceso que se empleó para elaborar este modelo de sobrevivencia, que incluye un análisis de datos (de forma univariada y bivariada), posteriormente se referencian los supuestos del modelo y se muestra el modelo para la deserción y en seguida el de graduación, en donde para cada uno, se emplean el estimador de la función de sobrevivencia Kaplan-Meier, función de riesgo acumulada, adicionado a esto se seleccionan las variables, luego se estima el modelo de regresión de Cox , y por último se validan los supuestos de este.

6.1. DESCRIPCIÓN DE LOS DATOS

Esta sección muestra, en primer lugar, una descripción de forma univariada para cada una de las variables de interés. Algunas de ellas se refieren a aspectos personales, familiares, académicos e institucionales. De cada una se muestran algunos estadísticos como: las media, la mediana, mínimo y máximo. En segundo lugar se hace el análisis bivariado a través de tablas de contingencia, pruebas de correlación de Pearson.

La tabla 28 muestra el resumen numérico y descriptivo de las variables objeto de estudio.

Tabla 28 Resúmenes numéricos de las variables para ATH

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Género	F= Mujer	118 (68,2 %)
	M= Hombre	55 (31,8 %)
Nivel educativo de la madre	Básica Primaria	61 (35,26 %)
	Básica Secundaria	60 (34,68 %)
	Técnico	18 (10,4 %)
	Universitario	11 (6,35 %)
	Sin Clasificar	23 (13,29 %)

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Número de Hermanos	1 = Uno 2= Dos 3 = Tres 4 = Cuatro 5 = Cinco 6 = Seis Sin Clasificar = SC	16 (9,24 %) 77 (44,5 %) 40 (23,12 %) 9 (5,2 %) 6 (3,46 %) 2 (1,15 %) 24 (13,87 %) Media = 2.453 Min – Max = 1 – 6 Mediana = 2
Edad de Ingreso al programa	En Años	Media = 19,06 Min – Max = 16 – 40 Mediana = 19
Ingreso Familiar al Presentar el Examen de Estado	Bajo Medio Alto Sin Clasificar= SC	135(78,03%) 3 (1,73 %) 8 (4,62%) 27 (15,60 %)
Trabajaba al presentar el examen de estado	Si No Sin clasificar = SC	7 (4,1 %) 145 (83,8 %) 21 (12,1 %)
Vivienda Propia	Si No Sin Clasificar = SC	112 (64,739 %) 43(24,855 %) 18 (10,4 %)
Puntaje en el Examen de Estado	De 1 – 100 puntos	Media = 67,6 Min – Max = 3 - 98 Mediana = 71,5
Número de Materias Perdidas		Media = 8,067 Min – Max = 1 - 31 Mediana = 7
Número de Materias Repetidas		Media = 8,06 Min – Max = 1 - 28 Mediana = 7
Tiempo Hasta la Deserción	Semestres Académicos (16 semanas)	Media =4,08 Min – Max = 1 - 13 Mediana =1
Estado del estudiante Dicotomizado	Desertor No Desertor	58 (33,52%) 115 (66,47 %)
Tipo de deserción	Académica No Académica	20(34,48%) 38(65,5%)

VARIABLE	UNIDADES DE MEDICIÓN /codificación	RESUMEN NUMÉRICO
Motivo de deserción Académica “Perdió Cupo/ Artículo 80”	Literal B	3 (15 %)
	Literal C	3 (15 %)
	Literal D	9 (45 %)
	Literal E	5 (25 %)
Tipo de Deserción	Precoz	21 (36,20 %)
	Temprana	20(34,48 %)
	Tardía	17 (29,31%)

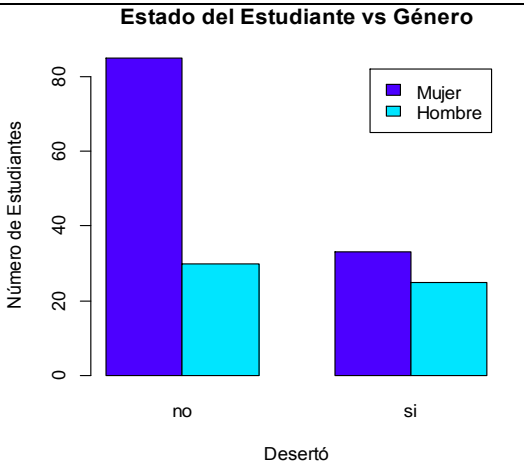
Fuente: Autor

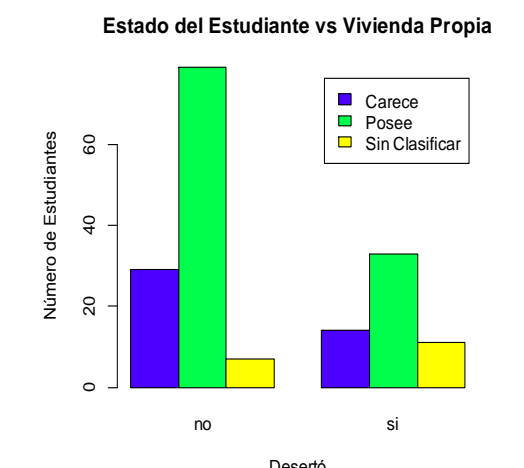
Con los resultados anteriores se puede observar que de los estudiantes seleccionados en la muestra, la mayoría son mujeres (68,2%). La edad media al momento de ingresar al programa fue de 19 años. El 65 % tenían vivienda propia y el 84 % no trabajaban al momento de presentar el examen de Estado. El puntaje promedio al momento de presentar el examen de estado fue de 67.6 en una escala de 0 a 100. El 78,03% tenían un nivel bajo de ingresos familiares al momento de presentar el examen de Estado. Para este programa, se tiene que a primer semestre de 2015, ha desertado el 33,52 % de los estudiantes de las cohortes entre el primer semestre de 2004 y el segundo semestre de 2009, presentado el 36,20 % deserción precoz.

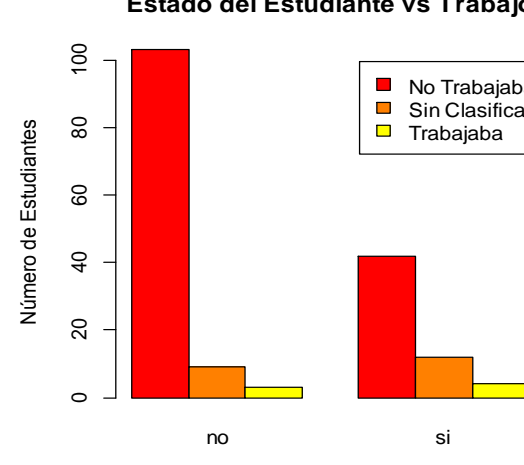
Además el 35,6 % de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

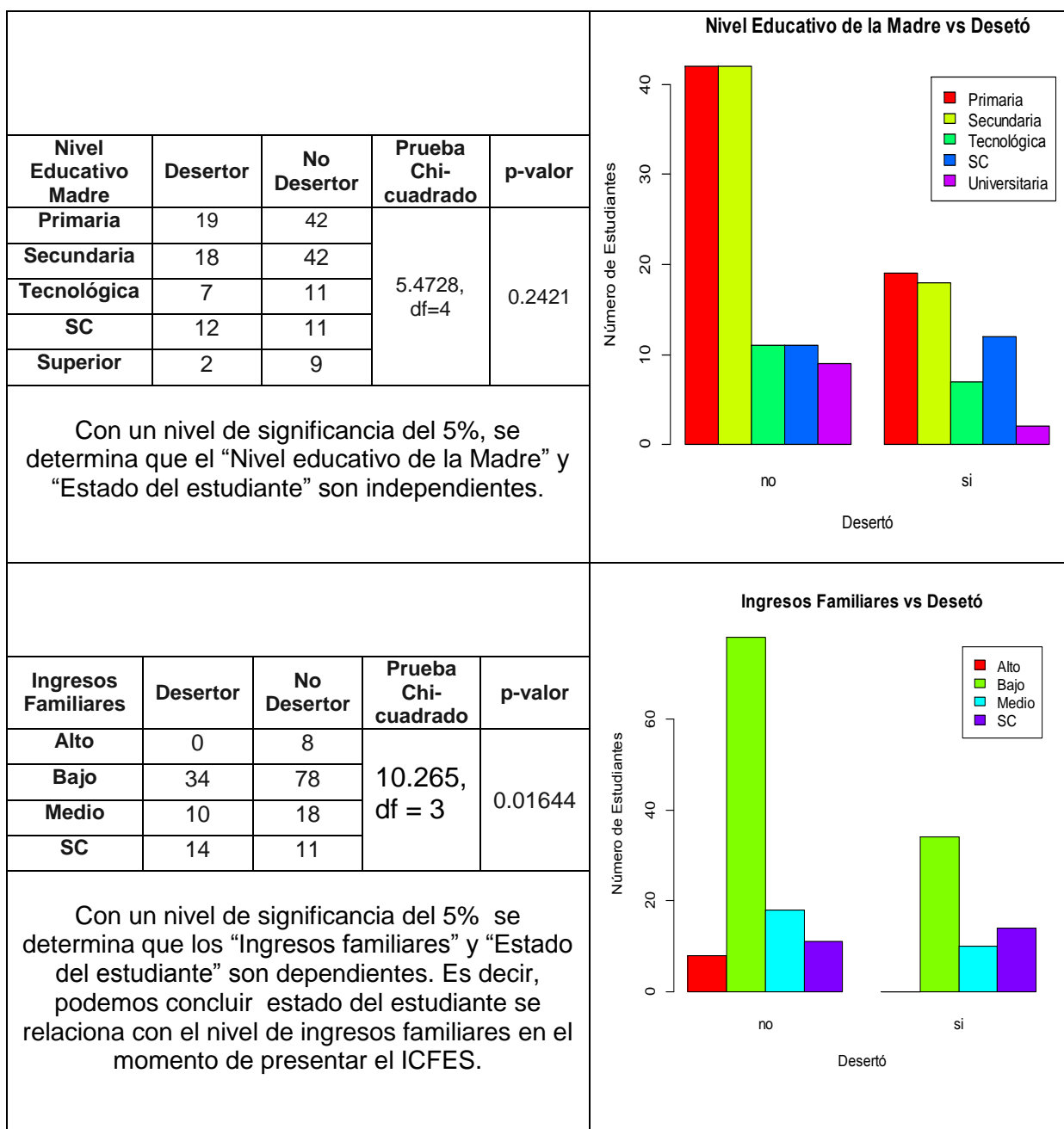
A continuación se presenta una descripción bivariada de las variables categóricas objeto de estudio y se pretende determinar si las variables categóricas se asocian con el estado del estudiante (desertor o no desertor), para lo cual se utilizó la prueba chi-cuadrado, además se hizo el diagrama de barras correspondiente a estas. En la tabla 29 se presenta lo encontrado.

Tabla 29 Contrastes de algunas variables categóricas del programa de ATH

					Estado del Estudiante vs Género	
Género	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor		
Hombre	25	30	5.1483, df=1	0.02327		
Mujer	33	85				
Con un nivel de significancia al 5%, se determina que el “Estado del Estudiante” y “Género” son dependientes. Es decir, podemos concluir estado del estudiante depende del género.						

					Estado del Estudiante vs Vivienda Propia	
Vivienda	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor		
Carece	14	29	6.9931 df=2	0.0303		
Posee	33	79				
SC	11	7				
Con un significancia del 5% las variables “Vivienda Propia” y “estado del estudiante” son dependientes, es decir que estas variables están relacionadas.						

					Estado del Estudiante vs Trabajo	
Trabajaba	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor		
Si	4	3	8.3608, df=2	0.01529		
No	42	103				
SC	12	9				
Hay evidencia estadística para determinar que el “Trabajaba en el momento de presentar el ICFES” y “Estado del estudiante” son dependientes. Es decir, se puede decir que el “desertar” depende de si trabajaba en el momento de presentar el ICFES.						



Fuente: Autor

Por último se realiza el test de correlaciones de Pearson entre el tiempo hasta la deserción y algunas variables cuantitativas continuas, para comprobar si el tiempo hasta la deserción de los estudiantes está correlacionada o no con algunas variables cuantitativas, los resultados se resumen en la tabla 30:

Tabla 30 Test de correlación de Pearson

Variable	Tiempo Hasta la Deserción(Semestres)		
	Coefficiente de Correlación	Estadístico t	p – valor
Edad de Ingreso al Programa	-0.1108643	-1.1539	0.2511
Número de Materias Perdidas	0.06023964	0.6243	0.5338
Número de Materias Repetidas	0.07154715	0.742	0.4597
Puntaje ICFES	0.1716810	1.8026	0.07426
Número de Hermanos	-0.002794405	-0.0289	0.977

Fuente: Autor

Con un nivel de significancia del 5% se encuentra que no hay evidencia estadística que relacione el tiempo en desertar con las variables mencionadas anteriormente.

6.2. MODELO DE SOBREVIDA PARA LA DESERCIÓN PARA ATH

Como el propósito del estudio es identificar los factores que están relacionados al tiempo de la deserción, n primer lugar, se presentará la aplicación de la metodología de Kaplan Meier para estimar la función de sobrevivida, con la cual se puede analizar la evolución de la probabilidad de deserción con su respectivo intervalo de confianza. Posteriormente, se construirá el modelo de regresión de Cox para estimar el efecto de las variables de estudio sobre los tiempos de sobrevivencia a la deserción. Las características del estudio son:

Tabla 31 Consideraciones para el Modelo de Sobrevida para ATH

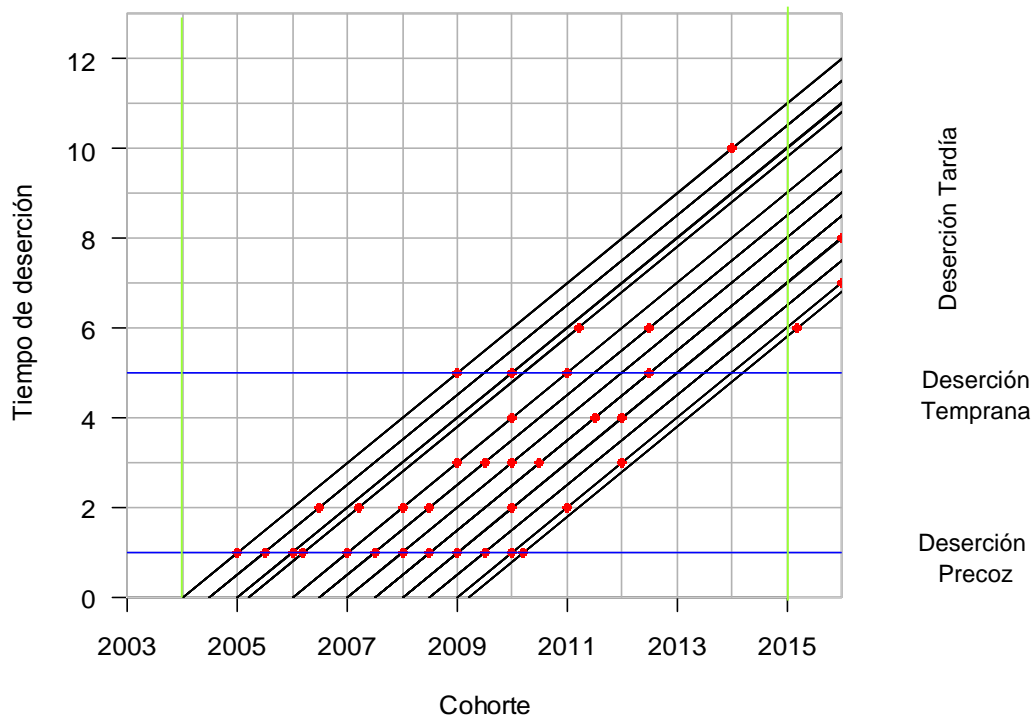
La unidad sobre la cual se registra el evento		Estudiantes seleccionados en la muestra, matriculados en cada programa de Administración turística y hotelera en primer semestre.
Evento de interés		Desertar del Programa
Variable respuesta		Tiempo hasta la Deserción: Cuya escala de medición es de razón de tipo discreto, ya que se mide en número de semestres cursados hasta que presenta el fenómeno de estudio.
Tiempo	Inicial del Estudio	I Semestre de 2004
	Origen del Evento	Cohorte del Estudiante (primera matricula del estudiante por programa)

	Final del Estudio	I Semestre de 2015
La escala de medida del tiempo hasta el evento		Semestres Calendario(16 Semanas)
Censura	Tipo	Tipo I (a derecha)
	Tiempo	13 semestres por cohorte

Fuente: Autor

A continuación se presenta la representación de los estudiantes del estudio mediante el diagrama de Lexis (ver sintaxis en el “Anexo E”) se evidencia que los estudiantes bajo estudio no tienen el mismo tiempo de origen, se señala con los puntos rojos aquellos que presentaron el evento y el tiempo en que sucedió (expresado en semestres). Nótese que la mayoría de estudiantes desertan precozmente, seguido de la deserción temprana, unos pocos se registran con deserción tardía.

Gráfica 37 Diagrama de Lexis – ATH

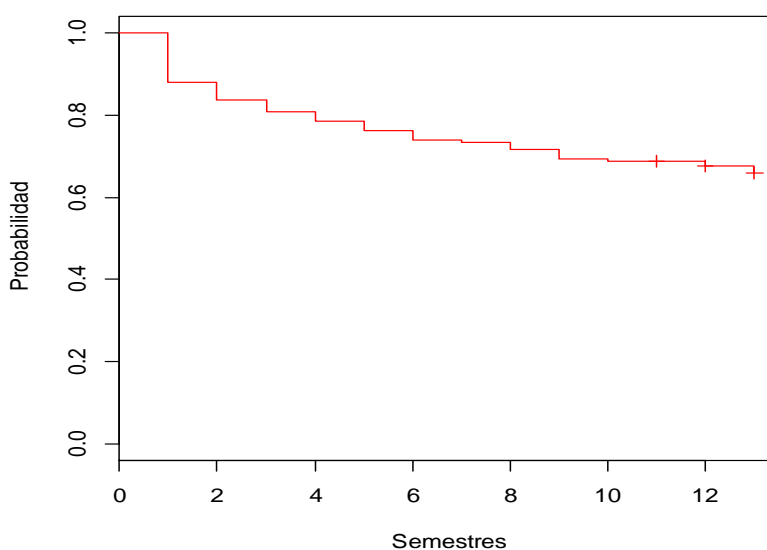


Fuente: Autor

6.2.1. Función de Sobrevida

La función de sobrevida es estimada por el método de Kaplan-Meier la cual se muestra a continuación. En el primer semestre se tiene una probabilidad de sobrevivir del 87.9%, en el cuarto semestre llega alrededor del 78.6% disminuyendo progresivamente, y como se puede observar en la gráfica la disminución es gradual a partir del tercer semestre y la probabilidad de no desertar es mucho mayor en el semestre 13 y es alrededor del 65.9%

Gráfica 38 Función de Sobrevida para ATH



Fuente: Autor

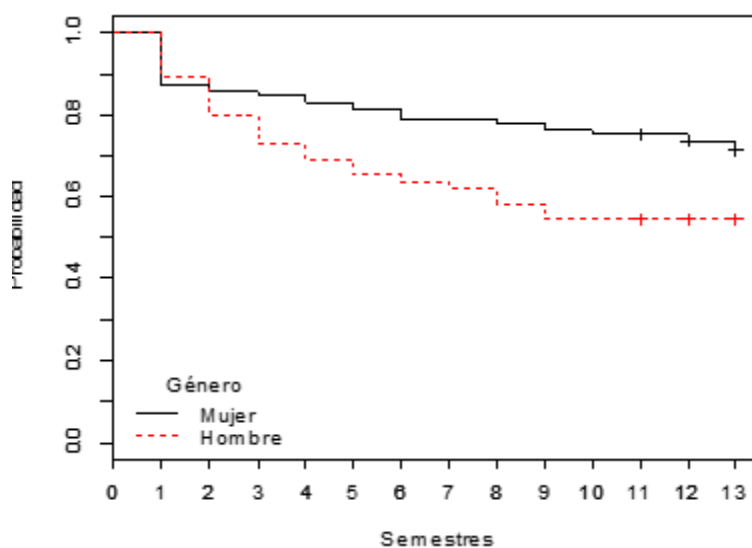
Tabla 32 Estimador de KM

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	173	21	0.879	0.0248	0.831	0.929
2	152	7	0.838	0.0280	0.785	0.895
3	145	5	0.809	0.0299	0.753	0.870
4	140	4	0.786	0.0312	0.727	0.850
5	136	4	0.763	0.0323	0.702	0.829
6	132	4	0.740	0.0334	0.677	0.808
7	128	1	0.734	0.0336	0.671	0.803
8	127	3	0.717	0.0343	0.653	0.787
9	124	4	0.694	0.0350	0.628	0.766
10	120	1	0.688	0.0352	0.622	0.760
12	108	2	0.675	0.0357	0.609	0.749
13	86	2	0.659	0.0366	0.592	0.735

Fuente: Autor

Teniendo en cuenta que el fenómeno de la deserción no se está estudiando de manera aislada, se hace necesario un análisis de la función de sobrevivencia separado por poblaciones de estudio, lo cual se presenta a continuación. En primer lugar, la función de sobrevivencia por género, de modo que en este tienen menos posibilidades sobrevivir a partir del segundo semestre los hombres y va disminuyendo esta gradualmente hasta el 9 semestre llegando a ser cercana al 50 %, por su parte las mujeres a partir del segundo semestre disminuye la posibilidad de continuar en el programa de una manera lenta, donde del 6 al 8 semestre, presentan igual probabilidad de permanecer, llegando así hasta un 70 % de culminar con sus proyectos educativos en el semestre 13 cerca de un 20% más en relación con los hombres en ese mismo semestre.

Gráfica 39 Función de Sobrevivencia por Género para ATH



Fuente: Autor

Al comparar las curvas de supervivencia de los hombres y las mujeres se obtiene un valor de $\chi^2 = 12.2$ con 1 grado de libertad y un p valor de 0.000469, estadística que nos permite afirmar que se puede rechazar la hipótesis nula de igualdad de las curvas de supervivencia. Por lo tanto podemos afirmar que la sobrevivencia (a la deserción) de hombres y mujeres no tiene el mismo comportamiento.

Tabla 33 Estimador de KM por género

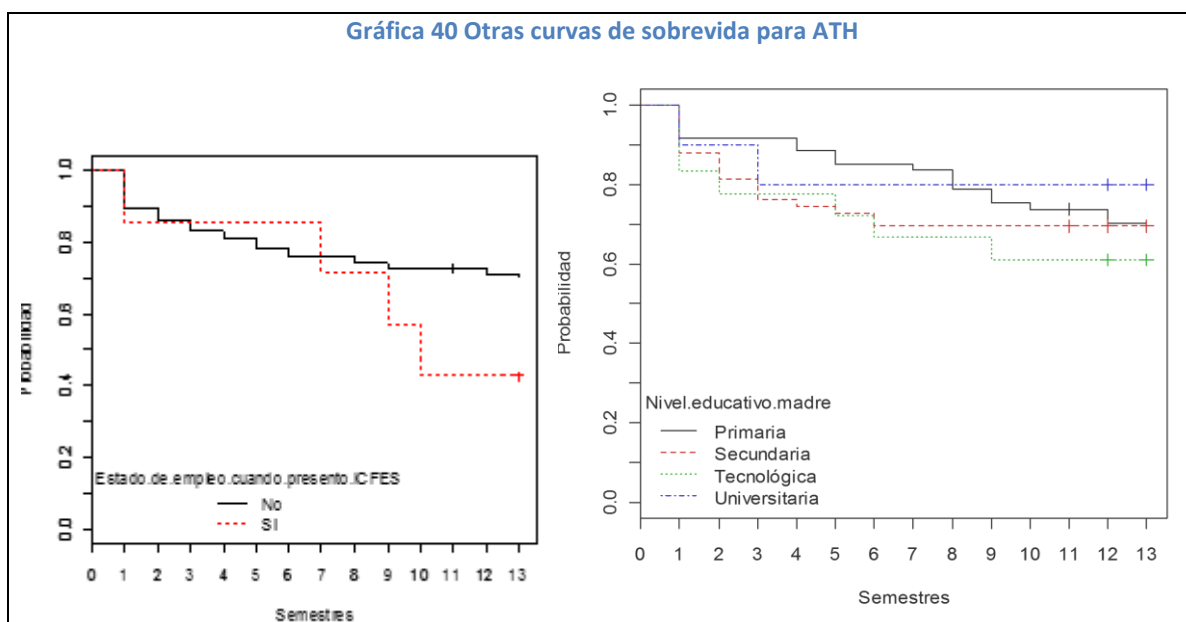
Género=Femenino					95% CI		Género=Masculino					95% CI	
time	n.risk	n.event	survival	std.err	lower	upper	time	n.risk	n.event	Survival	std.err	Lower	upper
1	95	10	0.895	0.0315	0.835	0.959	1	52	10	0.808	0.0547	0.707	0.922
2	85	5	0.842	0.0374	0.772	0.919	2	42	8	0.654	0.0660	0.537	0.797
3	80	7	0.768	0.0433	0.688	0.858	3	34	7	0.519	0.0693	0.400	0.674
4	73	4	0.726	0.0457	0.642	0.822	4	27	2	0.481	0.0693	0.362	0.638
5	69	3	0.695	0.0472	0.608	0.794	5	25	3	0.423	0.0685	0.308	0.581
6	66	2	0.674	0.0481	0.586	0.775	7	22	2	0.385	0.0675	0.273	0.542
8	64	1	0.663	0.0485	0.575	0.765	8	20	1	0.365	0.0668	0.255	0.523
9	63	1	0.653	0.0489	0.564	0.756	10	19	1	0.346	0.0660	0.238	0.503
10	62	1	0.642	0.0492	0.553	0.746							
13	31	1	0.621	0.0518	0.528	0.732							

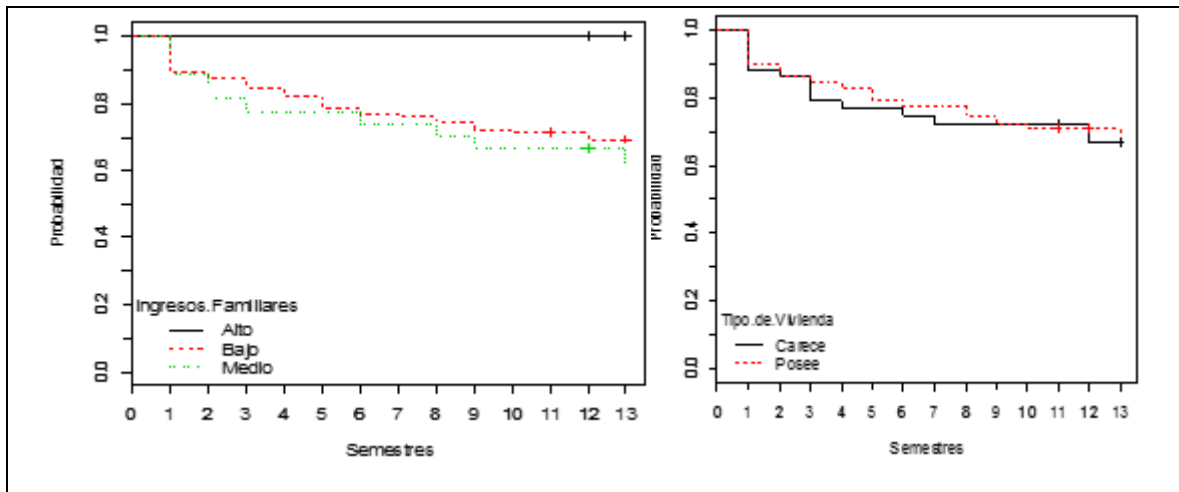
Fuente: Autor

Se efectuó la prueba de LongRank para juzgar la igualdad de las curvas de sobrevivencia por cada variable, como se muestra en como se muestra en la gráfica 40, encontrándose que para las diferentes categorías de variables no se rechaza la hipótesis nula de igualdad de estas, cómo se evidencia a continuación:

Variable	Chisq	Grados de Libertad	P valor
Nivel.educativo.madre	1.1	3	0.786
Tipo.de.Vivienda	0.1	1	0.732
Trabajaba.presento.ICFES	1.9	1	0.171
Ingresos.Familiares,	3.1	2	0.208

Gráfica 40 Otras curvas de sobrevida para ATH





Fuente: Autor

6.2.1. Factores asociados al riesgo de Deserción

En el apartado anterior lo que se hizo un análisis de la situación de sobrevivencia de los estudiantes del programa de ATH, con esto se pretende llegar a explicar los factores relacionados con el riesgo de deserción en los estudiantes mediante el modelo de regresión de Cox. Inicialmente se presenta el modelo completo con todas las variables, luego para la obtención del mejor conjunto de variables explicativas del modelo se tiene en cuenta con el método de selección hacia adelante con el Criterio de Información de Akaike (AIC). Posteriormente se calcula la curva de Sobrevivencia. Las variables a considerar son el género, edad de ingreso al programa, nivel de ingresos familiares, nivel educativo de la madre, puntaje estandarizado del ICFES, estado de empleo cuando presentó el ICFES, tenencia de vivienda y número de hermanos.

El modelo de regresión de Cox inicial con todas las covariables, se presenta a continuación:

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	6.728e-02	1.070e+00	3.922e-02	1.715	0.0863 .
Trabajaba.cuando.presento.ICFES[T.Si]	4.573e-01	1.580e+00	6.558e-01	0.697	0.4856
Género[T.M]	9.628e-01	2.619e+00	3.771e-01	2.554	0.0107 *
Ingresos.Familiares[T.Bajo]	1.743e+01	3.707e+07	3.708e+03	0.005	0.9963
Ingresos.Familiares[T.Medio]	1.774e+01	5.085e+07	3.708e+03	0.005	0.9962
Nivel.educativo.madre[T.Secundaria]	1.618e-01	1.176e+00	4.835e-01	0.335	0.7378
Nivel.educativo.madre[T.Tecnológica]	9.318e-02	1.098e+00	6.580e-01	0.142	0.8874
Nivel.educativo.madre[T.Universitaria]	9.923e-02	1.104e+00	1.156e+00	0.086	0.9316
No..De.hermanos	-1.066e-01	8.989e-01	1.878e-01	-0.568	0.5703
Puntaje.estandarizado.ICFES	1.878e-02	9.814e-01	7.946e-03	-2.364	0.0181 *
Tipo.de.Vivienda[T.Posee]	1.398e-01	1.150e+00	5.094e-01	0.275	0.7837

Seguidamente se procede a seleccionar el modelo más parsimonioso, utilizando el método de selección hacia adelante y el criterio AIC(en el Anexo D se muestra en detalle esta selección), encontrándose que el modelo óptimo queda determinado por el género, puntaje del ICFES estandarizado, edad de ingreso al programa. Se muestra el modelo de Cox:

	coef	exp(coef)	se(coef)	z	Pr(> z)
Género[T.M]	9.200e-01	2.509e+00	3.566e-01	2.580	0.00988 **
Puntaje.estandarizado.ICFES	-1.836e-02	9.818e-01	7.766e-03	-2.365	0.01804 *
Ingresos.Familiares[T.Bajo]	1.716e+01	2.827e+07	3.704e+03	0.005	0.99630
Ingresos.Familiares[T.Medio]	1.762e+01	4.479e+07	3.704e+03	0.005	0.99621
Edad.de.ingreso.al.programa	6.960e-02	1.072e+00	3.648e-02	1.908	0.05642 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Género[T.M]	2.509e+00	3.985e-01	1.2474	5.0475
Puntaje.estandarizado.ICFES	9.818e-01	1.019e+00	0.9670	0.9969
Ingresos.Familiares[T.Bajo]	2.827e+07	3.537e-08	0.0000	Inf
Ingresos.Familiares[T.Medio]	4.479e+07	2.233e-08	0.0000	Inf
Edad.de.ingreso.al.programa	1.072e+00	9.328e-01	0.9981	1.1515

En cuanto a las características observadas en los estudiantes se tiene que el género, puntaje en el ICFES y la edad de ingreso al programa, son significativas en el momento de explicar la deserción en el programa de ATH.

Lo que quiere decir que los Hombres tienen un riesgo de desertar de 2.509 mayor que las mujeres, hecho que es consistente con otros estudios y con los resultados de los anteriores programas, (Giovagnoli, 2002) afirma que las mujeres poseen mayor responsabilidad y perseverancia de ello mayor probabilidad de sobrevivir a lo largo del tiempo.

Los resultados también indican que el cambio proporcional en la función de riesgo que resulta de un aumento en un punto del puntaje estandarizado del ICFES con el que un estudiante ingresa al Programa, es negativo, lo que significa que el puntaje del ICFES incide en el riesgo de desertar, el cual disminuye a medida que le estudiante tiene mayor puntaje. La estimación del *riesgo relativo* del puntaje estandarizado en el ICFES es 0.9818, con un intervalo de confianza al 95% (0.9670, 0.9969) que indica que si comparamos dos estudiantes, manteniendo las demás variables constantes, para aquel estudiante que tenga un punto más en la prueba, se disminuye la probabilidad de desertar en 1.019. Este es un resultado consistente en el entendido que la prueba es una buena medida de las habilidades y conocimientos, los cuales son fundamentales en los primeros semestres del programa ya que éstos influyen en el resultado académico y consecuentemente en su riesgo de desertar.

En cuanto al nivel de ingresos del estudiante si es bajo tiene un riesgo muy alto en desertar con respecto a los que tienen ingresos altos. Además si este ingreso es medio se tiene un riesgo mayor de presentar el evento con respecto al de ingresos altos. Se observa que esta variable no es estadísticamente significativa pero su combinación con las otras resulta relevante.

Las variables que aumentan el riesgo en desertar son ser hombre, tener ingresos familiares en nivel medio y bajo, por el contrario disminuyen esta posibilidad el ingresar con un mayor puntaje en el ICFES.

En cuanto a la bondad de ajuste del modelo es aceptable para cualquiera de los tres criterios, ya que indican que el vector de parámetros es distinto de cero con un nivel de significancia del 5%, además las variables seleccionadas para el estudio explican el 16.2% de la variabilidad en el tiempo de deserción.

Tabla 34. Prueba de Verosimilitud, Wald y Score.

Rsquare= 0.162	(max possible= 0.94)		
Likelihood ratio test= 19.26	on 5 df,	p=0.001721	
Wald test	= 15.21 on 5 df,	p=0.009504	
Score (logrank) test = 19.14	on 5 df,	p=0.001812	

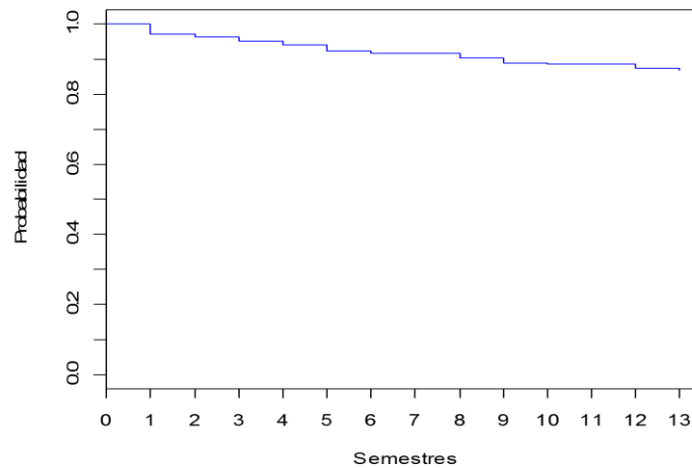
Se muestra en la gráfica 41 y la tabla 35, la función de sobrevida para el modelo de Cox en Deserción estimado anteriormente, la cual no presenta variaciones a la función de sobrevivencia estimadas sin considerar las variables, así:

Tabla 35 Función de Sobrevida Modelo de Deserción en ATH

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	109	9	0.973	5.44	1.70e-05	1
2	100	3	0.963	7.36	2.99e-07	1
3	97	3	0.952	9.52	2.95e-09	1
4	94	3	0.940	11.77	2.11e-11	1
5	91	4	0.924	14.82	2.07e-14	1
6	87	2	0.916	16.39	5.44e-16	1
8	85	3	0.903	18.76	1.89e-18	1
9	82	3	0.890	21.20	4.59e-21	1
10	79	1	0.885	22.03	5.81e-22	1
12	75	2	0.875	23.80	6.21e-24	1
13	61	1	0.869	24.87	3.77e-25	1

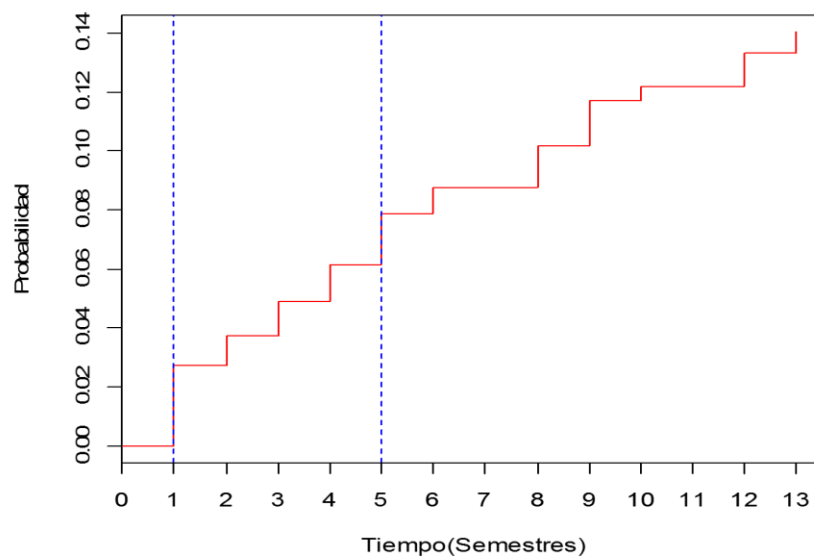
Fuente: Autor

Gráfica 41 Función de sobrevivencia Modelo de Deserción en ATH



Con base en la función de riesgo acumulada se puede afirmar que la probabilidad que un estudiante deserte justo al terminar el primer semestre es apenas del 3% es decir que hay una probabilidad alta de que no deserte precozmente del Programa. De igual manera se observa que la función de riesgo presenta cambios en la deserción temprana, pero estos no son grandes por ejemplo, se puede apreciar, que parece estar casi constante el riesgo a partir del segundo que es alrededor de 3.7% consecuentemente en cada semestre aumenta en 1%.

Gráfica 42 Función de riesgo acumulado Modelo de Deserción de ATH



Fuente: Autor

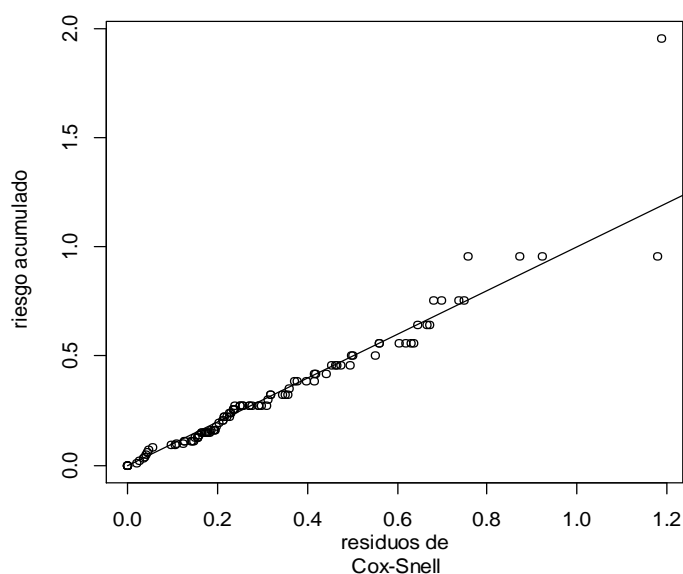
6.2.3. Evaluación del modelo

Al verificar el supuesto de riesgos proporcionales del ajuste del modelo de regresión para ATH se obtiene que:

	rho	chisq	p
Género[T.M]	-0.0257	2.18e-02	0.883
Puntaje.estandarizado.ICFES	-0.1504	9.03e-01	0.342
Ingresos.Familiares[T.Bajo]	0.2255	3.15e-09	1.000
Ingresos.Familiares[T.Medio]	0.1911	7.99e-09	1.000
Edad.de.ingreso.al.programa	0.0895	1.41e-01	0.707
GLOBAL		1.18e+00	0.947

Por lo que, no existen evidencias significativas al 5% de que se esté violando el supuesto de riesgos proporcionales para las variables incluidas en el modelo ni de manera global. Luego este supuesto se aborda de manera gráfica mediante los residuos de Cox-Snell, si el modelo es correcto y la estimación de los β son cercanas a los valores reales, entonces este estimador debería describir aproximadamente una recta de pendiente igual a la unidad. La gráfica 43 indica que, en general, este modelo se ajusta a los datos, excepto en algunos donde las estimaciones son inestables debido a la censura de los datos.

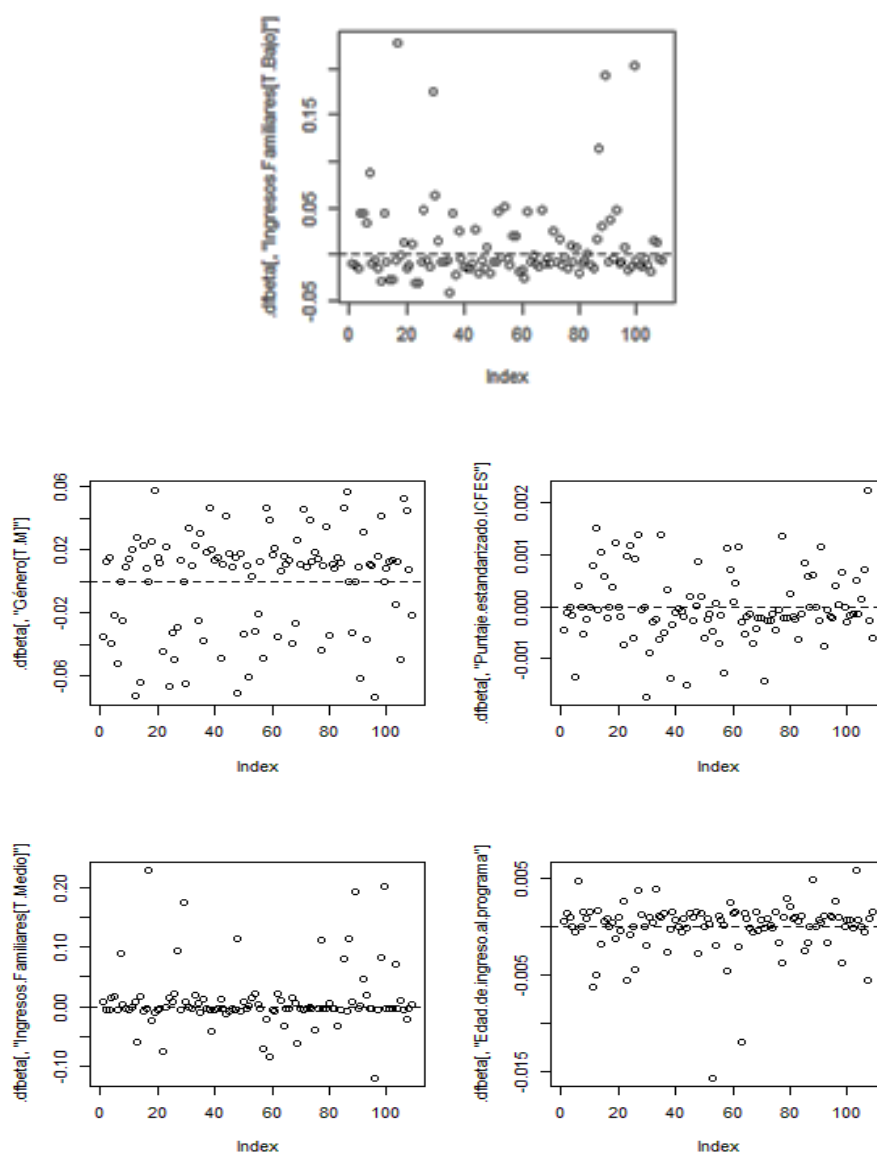
Gráfica 43 Residuos de Cox-Snell



Fuente: Autor

Comprobación de la hipótesis de riesgos proporcionales por covariables: Residuos escalados de Schoenfeld: Para determinar la influencia de cada observación en el modelo ajustado, se representan en la gráfica 44 los residuos $dfbeta$. Es decir, para la cada covariable se representa la observación por el cambio en cada semestre (dividiendo por el error estándar) del coeficiente después de la eliminación de la observación del modelo, además una gráfica de β correcta, frente al tiempo describirá una línea horizontal.

Gráfica 44 Residuos escalados de Schoenfeld



Fuente: Autor

Todas las covariables presentan una concentración alrededor de cero por tanto representan una gráfica de β correcta frente al tiempo.

6.2.4. Pronóstico a partir del modelo

Como en la sección 2.4.5 se mencionó que la función de riesgo inicial $h_0(t)$ depende del tiempo, mientras que $\exp(X\beta^T)$ depende únicamente de las covariables. Sin embargo, en el modelo de riesgos proporcionales de Cox se debe tener en cuenta que los riesgos para dos conjuntos diferentes de valores de los covariables conservan la misma proporción a lo largo del tiempo. Con base en los resultados obtenidos en el modelo de deserción en ATH, donde este corresponde a:

$$\begin{aligned} h_i(t, X, \beta) = h_0(t) \exp & \left((0.920 * \text{Género}[M]) + (0.0186 * \text{Punt. ICFES}) \right. \\ & + (17.36 * \text{Ingre. Fami}[T. Bajo]) + (17.62 * \text{Ingre. Fami}[T. Medio]) \\ & \left. + (0.0692 * \text{Edad. Ingreso}) \right) \end{aligned} \quad (5.26)$$

El exponente del modelo de Cox, que particularmente para el estudio corresponde a la ecuación (5.26) se denomina *índice pronóstico* (P_I), (Domnéch, 1992). De manera que la tasa de riesgo se puede expresar a través de este indicador en lugar utilizar los valores del estudiante para cada variable. En la práctica es mejor usar el *índice de pronóstico centrado* (P_{IC}), que se obtiene centrando las variables predictoras:

$$P_{IC} = \beta_1(X_1 - \bar{X}_1) + \dots + \beta_p(X_p - \bar{X}_p)$$

Que para el programa de AI, corresponde a:

$$\begin{aligned} P_{IC} = & (0.920 * \text{Géner}[M]) + (0.0186 *)(X_1 - 53,29) + (17.36 * \text{Ingre. Fami}[T. Bajo]) + \\ & (17.62 * \text{Ingre. Fami}[T. Medio]) + (0.0692)(X_1 - 19.02) \end{aligned}$$

Un índice pronóstico centrado igual a cero para el programa de ATH, corresponde a una estudiante mujer con edad promedio de 19.02 años y puntaje estandarizado

ICFES de 53,29 con ingresos familiares altos, y que se denomina de ahora en adelante “estudiante base”. La diferencia entre los índices pronósticos de dos estudiantes permite estimar su riesgo relativo (RR), veamos:

$$RR = \frac{h_i(t, X_i, \beta)}{h_j(t, X_j, \beta)} = \frac{h_0(t) \exp(P_{IC\text{individuo}i})}{h_0(t) \exp(P_{IC\text{individuo}j})} = \exp(P_{IC\text{individuo}i} - P_{IC\text{individuo}j})$$

Teniendo en cuenta lo anterior a continuación se presenta el RR para cinco estudiantes, seleccionados aleatoriamente, y que se encuentran cursando el primer semestre del programa, es decir, estudiantes que ingresaron en el I semestre de 2015.

Tabla 36 Pronóstico del modelo de deserción para AI

N°	Estudiante				P_{IC}	RR respecto a un estudiante promedio
	Género	Edad.Ingr	Puntaje ICFES	Nivel de Ingresos Familiares		
1	F	16	87	Bajo	1,409216	4,09274543
2	M	19	88	Bajo	1,27088	3,56398749
3	F	20	84	Bajo	1,375016	3,95514
4	F	18	76	Bajo	1,343016	3,83057913
5	M	19	85	Bajo	1,4729	4,36186623

Lo anterior significa que el riesgo de deserción para todos los estudiantes seleccionados aleatoriamente, resultan tener un RR alto.

6.3. MODELO PARA LA GRADUACIÓN DE ATH

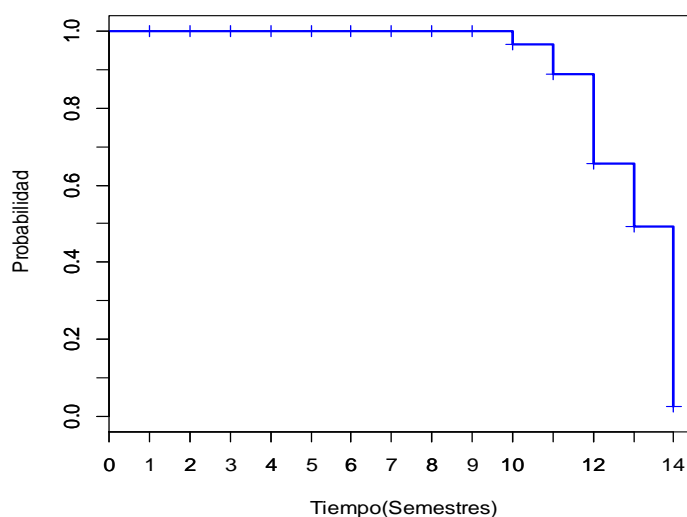
La duración teórica de la carrera universitaria es de cinco años. Sin embargo, los alumnos que no desertan prolongan sus estudios por un mayor tiempo. Este apartado se concentra en una pregunta: ¿Qué factores “favorecen” a la finalización exitosa del programa de AI dado el tiempo que lleva el alumno estudiando? En este

sentido se sigue presentando censura tipo I a derecha, y el tiempo de censura se define como el percentil 75% del tiempo de permanecía los estudiantes en el programa y corresponde a 14 semestres.

6.3.1. Función de Sobrevida

En la curva que muestra la gráfica 45 se aprecia, que la probabilidad de los estudiantes de llegar a décimo semestre y no se gradúen es de 96,3 %, de llegar a 11 y no graduarse es cercana al 88.74%, disminuyendo así en el semestre 13 llegando al 49,27% y en el semestre 14 la probabilidad de no graduarse es muy baja (1.7%).

Gráfica 45 Función de Sobrevida de Graduación para ATH



Fuente: Autor

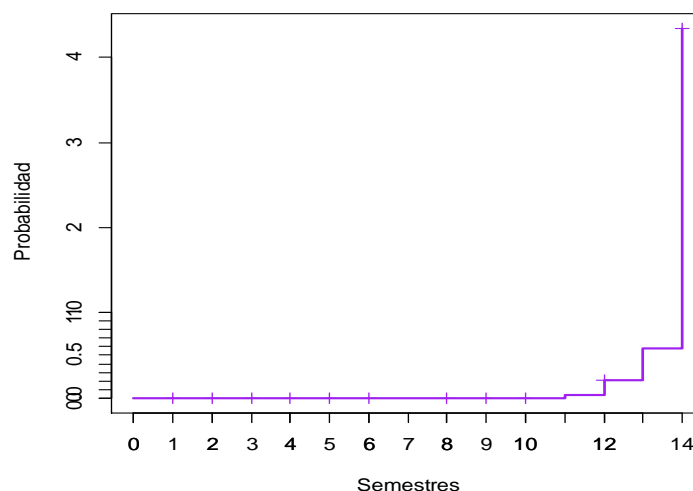
Tabla 38 Función de Sobrevida – Graduación

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
10	91	3	0.9670	0.0187	0.93104	1.0000
11	85	7	0.8874	0.0336	0.82399	0.9557
12	77	20	0.6569	0.0508	0.56447	0.7645
13	56	14	0.4927	0.0538	0.39770	0.6103
14	40	38	0.0246	0.0172	0.00627	0.0967

ATH Fuente: Autor

Además la probabilidad de terminar el décimo semestre y graduarse es 4.08% de llegar al semestre 13 y de graduarse en semestre catorce es muy alta, ya que la mayoría de eventos se sitúan en este semestre.

Gráfica 46 Función de riesgo para Graduación en ATH



Fuente: Autor

6.3.2. Factores asociados al aumento de Graduación

En seguida se presenta el modelo de regresión de Cox con todas las covariables:

n= 110, number of events= 76

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad.de.ingreso.al.programa	-0.057505	0.944117	0.030809	-1.867	0.062 .
Estado.de.empleo.cuando.presento.ICFES[T.Si]	0.588473	1.801235	0.735445	0.800	0.424
Género[T.M]	0.217479	1.242940	0.332380	0.654	0.513
Ingresos.Familiares[T.Bajo]	1.664020	5.280497	1.005549	1.655	0.098 .
Ingresos.Familiares[T.Medio]	1.453644	4.278677	0.945707	1.537	0.124
Nivel.educativo.madre[T.Secundaria]	0.087105	1.091012	0.335834	0.259	0.795
Nivel.educativo.madre[T.Tecnológica]	0.261316	1.298638	0.555409	0.470	0.638
Nivel.educativo.madre[T.Universitaria]	1.317341	3.733482	1.012480	1.301	0.193
No..De.hermanos	-0.094985	0.909386	0.153256	-0.620	0.535
Número.de.materias.repetidas.	-0.102961	0.902162	0.024537	-4.196	2.71e-05 ***
Tipo.de.Vivienda[T.Posee]	0.195777	1.216256	0.330661	0.592	0.554
X.Materias.perdidas.	0.004005	1.004013	0.018723	0.214	0.831

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

La selección del modelo a través del criterio AIC y con el método de selección hacia adelante, considera las variables de edad de ingreso al programa y el número de materias repetidas, de modo que:

```

              coef exp(coef) se(coef)      z Pr(>|z|)
Edad.de.ingreso.al.programa -0.04540  0.95562  0.02747 -1.653  0.0984 .
Número.de.materias.repetidas. -0.09602  0.90845  0.02277 -4.217 2.48e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Edad.de.ingreso.al.programa    0.9556      1.046    0.9055    1.0085
Número.de.materias.repetidas.    0.9084      1.101    0.8688    0.9499

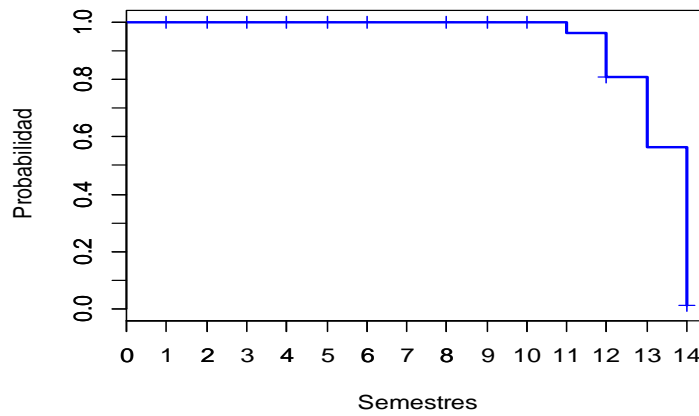
```

En cuanto a la edad de ingreso al programa, se tiene que el coeficiente del riesgo relativo es 0.96, con intervalo de confianza al 95% (0.905,1), como la estimación puntual es menor que 1, señala que si se comparan dos estudiantes, manteniendo las demás variables constantes, aquel que tenga un año más de edad se disminuye en 1.046 la probabilidad de graduarse.

Los resultados indican que riesgo relativo al número de materias repetidas es de 0.90845, con un intervalo de confianza al 95%(0.89, 0.97). Como la estimación puntual es inferior a 1, señala que si se comparan dos estudiantes, manteniendo las demás variables constantes, aquel estudiante que haya repetido una de más se disminuye en 1.101 la probabilidad de graduarse.

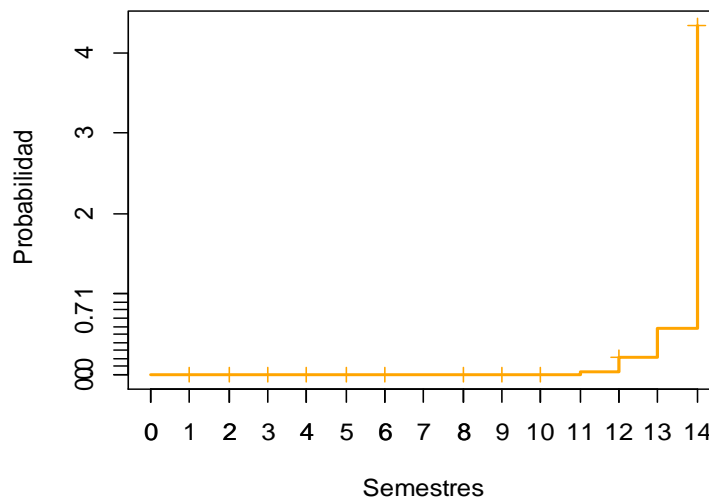
En seguida se presenta la función de sobrevivencia para el modelo estimado. Tal estimación se presenta en la tabla 39 y la curva de sobrevivencia para el modelo de Cox Ajustado de ATH en la gráfica 47.

Gráfica 47 Función de sobrevivencia del modelo para Graduación



Fuente: Autor

Gráfica 48 Función de riesgo acumulada modelo de Graduación



Fuente: Autor

Tabla 37 Modelo para Graduación en ATH

time	n.risk	n.event	survival	Std.err	lower95% CI	upper 95%CI
11	79	3	0.9620	0.0215	0.92079	1.000
12	76	12	0.8101	0.0441	0.72810	0.9014
13	62	19	0.5619	0.0564	0.46144	0.6841
14	43	42	0.0131	0.0130	0.00186	0.0916

Fuente: Autor

Para la bondad de ajuste del modelo, se aprecia un p-valor significativamente pequeño, lo cual es evidencia de que las covariable seleccionadas explicativas

tienen un coeficientes distintos de cero en el modelo de riesgos proporcionales, con un nivel de significancia del 5%, además el modelo tan explica solo el 14.4 % de la variabilidad al momento de graduarse un estudiante.

Tabla 38 Pruebas de Verosimilitud, Wald y Score.

Concordance= 0.823	(se = 0.076)
Rsquare= 0.174	(max possible= 0.991)
Likelihood ratio test= 21.09	on 2 df, p=2.633e-05
Wald test = 18.45	on 2 df, p=9.847e-05
Score (logrank) test = 18.92	on 2 df, p=7.784e-05

Fuente: Autor

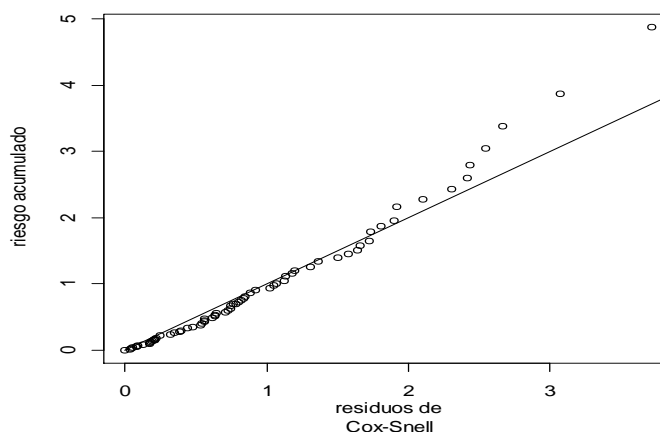
6.3.3. Evaluación del modelo

Al efectuar el Test de riesgos proporcionales con un nivel de significancia del 5%, se tiene que no se violan los estos supuestos para la covariable edad de ingreso al programa, pero sí de manera global, como se muestra enseguida:

	rho	chisq	p
Edad.de.ingreso.al.programa	0.149	1.46	0.227066
Número.de.materias.repetidas.	0.428	12.56	0.000393
GLOBAL		12.82	0.001647

Residuos de Cox-Snell: De manera general se comprueba la hipótesis de riesgos proporcionales, se mantienen ya que se ajustan a la línea recta, y en lo que concierne a la cola de la derecha se presenta por la censura de los datos para este programa.

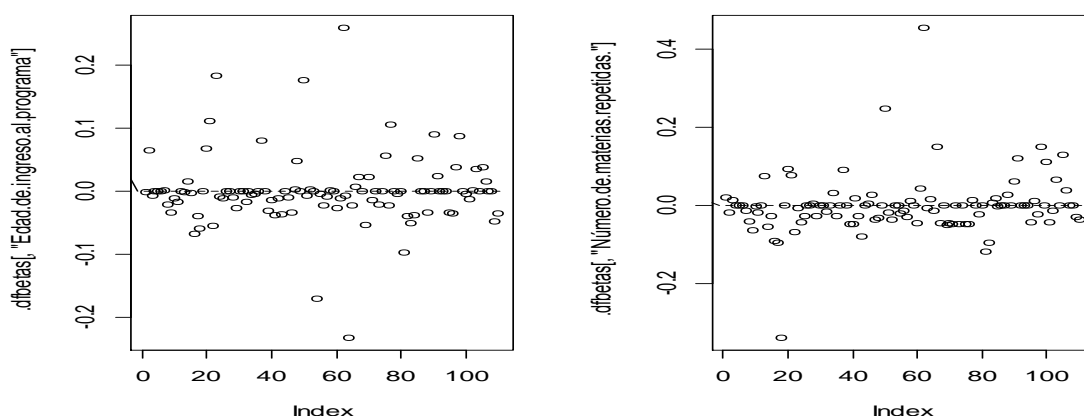
Gráfica 49 Residuos de Cox-Snell



Fuente: Autor

Residuos tipo dfBetas: Con este tipo de residuo, se determina la influencia de cada observación en el modelo. De manera que, para cada covariable, se ha representado la observación (en orden de tiempo de falla registrado) por el cambio de escala aproximada del coeficiente, es decir, dividiendo por el error estándar, después de la eliminación de la observación del modelo.

Gráfica 50 Residuos, tipo Df Betas



Fuente: Autor

En este caso ninguna de las observaciones ejerce un cambio mayor a 20% del error estándar como se observa en la gráfica 50. Si se comparan con el valor coeficiente de regresión estimado y su error estándar podemos concluir que ninguna de las observaciones es influyente individualmente sobre la estimación de los parámetros.

Como conclusión principal del programa de Administración turística y Hotelera se tiene que el 33,52% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 33, constituyéndose en una frustración personal, a nivel institucional las pérdidas

económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Desertan mayoritariamente los hombres cuya edad promedio de ingreso es de 19 años, además de los estudiantes que ingresan al programa el 65 % tenían vivienda propia y el 84 % no trabajaban al momento de presentar el examen de estado, el puntaje promedio al momento de presentar el examen de estado fue de 67.6 también el 35,6 % de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de ATH, la posibilidad de que un estudiante se vaya en primer semestre es del 13%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Donde las mujeres tienen una sobrevivencia superior a los hombres en un 20%.

Los factores que aumentan el riesgo de desertar hacen referencia a ser hombre, hecho que corrobora (Giovagnoli, 2002) donde lo atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar con sus proyectos educativos que la mujer, y tener ingresos familiares de nivel bajo o medio, ya que posiblemente a los requerimientos del programa se deba cumplir con prácticas inter-semestrales a diferentes lugares de la geografía colombiana o insumos de cocina o afines para la práctica de los conocimientos impartidos durante la carrera, se recomienda a la Universidad y el programa crear políticas de patrocinio y apoyo a aquellos estudiantes con este tipo de falacias económicas. Se disminuye el riesgo de desertar al presentar un alto puntaje en el ICFES en el momento de ingresar,

entendiendo que la prueba es una buena medida de las habilidades y conocimientos, los cuales son fundamentales durante el avance del programa ya que éstos influyen en el resultado académico.

Por el contrario, para la graduación los factores que disminuyen esta propensión son una mayor edad de ingreso al programa y que registre tener un mayor número de asignaturas repetidas, posee menor probabilidad de graduarse. Además el semestre en el que es más probable que un estudiante se vaya es el primero y se gradué en el 14. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Se sugirieren actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica de asignación de temas de grado en semestres inferiores o vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

CONCLUSIONES Y RECOMENDACIONES

1. Al comparar la sobrevivencia por programa, se encontró que Administración Turística y Hotelera tiene una sobrevivencia superior (entendida la sobrevivencia como la continuación en el Programa académico) y los que tienen menos probabilidades de sobrevivir son los estudiantes de Administración Industrial, en cuanto a Administración de Empresas Agropecuarias se encuentra en el promedio de los dos programas anteriores, y de manera general las mujeres tienen mejor nivel de sobrevivencia respecto a los hombres.

2. En el programa de Administración de Empresas Agropecuarias, se tiene que la deserción es del 46%, cifra que según el reporte del MEN (2014) supera la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2%, sin embargo en cuanto a la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines es del 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 46, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría (46,45%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Pero también es considerable la proporción de aquellos que lo hacen por causas no académicas predominado aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan de igual forma los hombres y mujeres, en el momento de presentar el ICFES no trabajaban, que además reportan no tener vivienda propia, cuya madre alcanza el nivel educativo de primaria, y sus ingresos familiares son de nivel bajo. El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, es decir del segundo al quinto semestre. Este fenómeno se

atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Los factores que aumentan el riesgo de desertar son los de ser hombre y el de tener ingresos en nivel medio, hecho que es consistente con lo planteado por (Giovagnoli, 2002) quien lo atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa y al registrar ingresos familiares en nivel bajo, hecho particular que refleja que este tipo de estudiantes son mucho más comprometidos al hacer un esfuerzo mayor de avanzar dentro del Programa, posiblemente abarcando las oportunidades que ofrece la Universidad. Con base en lo anterior, la Universidad y el Programa en particular deberían diseñar acciones preventivas de acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración de Empresas agropecuarias es de 20 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas y tener un puntaje alto en el ICFES disminuyen la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. Por el contrario registrar ingresos familiares en nivel bajo o medio, aumentan esta probabilidad de graduarse.

3. En cuanto al programa de Administración Industrial, se tiene que la deserción acumulada está cerca del 50%, cifra que supera según el reporte del MEN (2014) la del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 50, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría (59.6%) desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Sin embargo, también es considerable la proporción (40.4%) de los estudiantes que salen por causas académicas predominando aquellos que pierden una asignatura en calidad de repitente siendo su promedio académico acumulado inferior a 3.0. Desertan mayoritariamente los que en el momento de presentar el ICFES no trabajaban y que además reportan tener vivienda propia, cuya madre alcanza el nivel educativo de secundaria, y sus ingresos familiares son de nivel medio.

El periodo crítico en donde se presenta con mayor intensidad es de manera temprana, entre el segundo y quinto semestre. Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Se tiene que a partir de la culminación del primer semestre hay una marcada y significativa diferencia en la probabilidad de continuar con el Programa siendo la de las mujeres superior a la de los hombres.

Al observar los factores que explican el hecho de desertar se encontró que aumenta el riesgo de desertar en ser hombre, hecho que va de la mano con lo que sucede a nivel Nacional. Una explicación a este hecho es la dada por Giovangnoli, 2002 en donde se le atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar una carrera que la mujer. El riesgo de desertar disminuye con una mayor edad de ingreso al Programa. Con base en ese hecho la Universidad y el Programa en particular deberían diseñar acciones preventivas de acompañamiento académico ya que, como se observó, la edad promedio de ingreso a Administración Industrial es de 19 años.

Para el caso de la graduación, como era de esperarse, el número de materias repetidas disminuye la probabilidad de graduarse. La mayor probabilidad de graduarse se da en el semestre 13. Hecho que es preocupante ya que gasta tres semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil insertarse a la vida laboral. Este hecho podría ser menguado asignando temas de grado, para desarrollar monografías, desde el séptimo semestre o con la vinculación inmediata a algún programa de posgrado, según la normatividad de la Universidad. En este punto valdría la pena hacer un análisis de los posgrados que ofrece la Facultad y que sean afines al área de administración.

4. En el programa de Administración turística y Hotelera se tiene que el 33,52% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa - Economía, Administración, Contaduría y Afines - que se encuentra en 46.2%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 33, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Desertan mayoritariamente los hombres cuya edad promedio de ingreso es de 19 años, además de los estudiantes que ingresan al programa el 65 % tenían vivienda propia y el 84 % no trabajaban al momento de presentar el examen de estado, el puntaje promedio al momento de presentar el examen de estado fue de 67.6 también el 35,6 % de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5 % de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de ATH, la posibilidad de que un estudiante se vaya en primer semestre es del 13%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas. Donde las mujeres tienen una sobrevivencia superior a los hombres en un 20%.

Los factores que aumentan el riesgo de desertar hacen referencia a ser hombre, hecho que corrobora (Giovagnoli, 2002) donde lo atribuye a factores culturales o bien a que el hombre es menos perseverante al continuar con sus proyectos educativos que la mujer, y tener ingresos familiares de nivel bajo o medio, ya que posiblemente a los requerimientos del programa se deba cumplir con prácticas inter-semestrales a diferentes lugares de la geografía colombiana o insumos de cocina o afines para la práctica de los conocimientos impartidos durante la carrera, se recomienda a la Universidad y el programa crear políticas de patrocinio y apoyo a aquellos estudiantes con este tipo de falacias económicas. Se disminuye el riesgo de desertar al presentar un alto puntaje en el ICFES en el momento de ingresar, entendiendo que la prueba es una buena medida de las habilidades y conocimientos, los cuales son fundamentales durante el avance del programa ya que éstos influyen en el resultado académico.

Por el contrario, para la graduación los factores que disminuyen esta propensión son una mayor edad de ingreso al programa y que registre tener un mayor número de asignaturas repetidas, posee menor probabilidad de graduarse. Además el semestre en el que es más probable que un estudiante se vaya es el primero y se gradué en el 14. La mayor probabilidad de graduarse se da en el semestre 14. Hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Se sugirieren actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica de asignación de temas de grado en semestres inferiores o vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

5. A partir de los resultados obtenidos se puede apreciar que los factores estudiados que afectan la deserción y la graduación en cada una de las Administraciones de la Facultad seccional Duitama (excepto el número de materias repetidas) son exógenos, difícilmente pueden ser cambiados por la institución a la hora de implementar estrategias y políticas para disminuirla. Los riesgos se asocian más con las características de la población estudiantil de la Facultad. De modo que, es importante que la UPTC encamine sus esfuerzos a mecanismos alternativos de estudio que les permita a estudiantes con ciertas características “desfavorables” reducir los riesgos de deserción y aumentar sus riesgos de graduación

6. El análisis de sobrevivencia es una potente herramienta, muy importante que ha venido cobrando fuerza no solo en el área médica, por la inclusión de la censura y otras características, sino en otros campos del saber en dónde permite a través del modelamiento establecer cuáles son los factores asociados a cierto fenómeno referidos al tiempo hasta la ocurrencia de un evento, en este caso permitió establecer los factores asociados a la deserción y graduación universitaria en algunos programas de la Facultad Seccional Duitama.

7. Se recomienda construir a futuro nuevos modelos de sobrevivencia a partir de la recolección de información para cada programa administración de la Facultad Seccional Duitama, que contenga otros aspectos importantes y no medidos en este proyecto, para tal efecto este trabajo deja indicadas algunas variables y el formato de Excel para el registro, ver “Anexo H”.

8. Para estudios posteriores, se recomienda modelar el fenómeno de la deserción y graduación universitaria, de forma paramétrica asumiendo sobre el tiempo de deserción o graduación alguna distribución de probabilidad.

BIBLIOGRAFÍA

Aguilar, Angel (2009). Introducción al Análisis de Supervivencia con R. México.

Barrera, Mauricio. (2008). Análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira. Proyecto de maestría.

Borges, R. (2005). Análisis de Supervivencia Utilizando el Lenguaje R. Paipa, Boyacá, Colombia.

COLOSIMO, Enricon A. y GIOLO Ruíz, Suely.(2006)Conceptos básicos y ejemplos, técnicas no paramétricas y modelos probabilísticos. En: *Análise de sobrevivência aplicada*. 1 Ed. Edição: 1ª . EDGARD BLUCHER, 1997. P. 6-64.

Cox, D.R(1972) "Regression models and life tables" en Journal of the Royal Statistic Society, series B, vol 34, num 2, pp 187 - 220.

Doménech M, José (1992)" Una aplicación del analisis de la supervivencia en ciencias de la salud "." (Universitat Autònoma de Barcelona). Anuario de psicología, Núm. 55 (1992) , p. 109-141, ISSN 0066-5126.

García, C. P. (2014). Estudio de análisis de supervivencia.

Giovangnoli, P. I. (2002). Determinantes de la deserción y graduación universitaria: Una Aplicación de los modelos de duración.Documento de trabajo , núm 37, Argentina, Universidad Nacional de la Plata.

Godoy, A. M. (2009). Introducción al Análisis de Supervivencia con R. Universidad Autonoma de México.

HIGUITA RICO, Alberto D. *Caracterización de la deserción estudiantil en la Universidad Nacional de Colombia sede Medellín*. Medellín: Colombia. Universidad Nacional de Colombia sede Medellín. 2006. p. 8-13

Lopera, C.(2008). "Determinantes de la deserción Universitaria en la facultad de economía" Universidad del Rosario, en borradores de Investigación, pp. 1 -25.

Lerma, H. D. (2009). *Metodología de la Investigación: Propuesta, Anteproyecto, Proyecto*. Bogotá: Eco Ediciones .

Lininger, C. A., & Warwick, D. P. (1978). *La Encuesta por Muestreo: Teoría y Práctica*. México: Continental, S.A .

Ministerio de Educación Nacional (2009). "Deserción Estudiantil en la Educación Superior Colombiana: Metodología de seguimiento, diagnóstico y elementos para su prevención". Obtenido de:

http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf. [Recuperado el 25 de Febrero de 2015]

Ministerio de Educación Nacional (2014). Caracterización de estudiantes y Diagnóstico de deserción en educación Superior, en: http://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-254702_archivo_pdf_estadisticas_2013.pdf , recuperado el 18 de Febrero 2016.

Ospina, D. B. (2001). *Introducción al Muestreo*. Bogotá.D.C: UNIBIBLOS.

Pérez, C. (2000). *Técnicas de Muestreo Estadístico*. Madrid: Alfaomega.

SMITH, Peter J. (2002) Modelo de riesgo proporcional de Cox. En: *Analysis of failure and Survival data* . Chapman&Hall, 1997. P. 167-181.

Osorio, Ana-María, Catalina Bolancé y Maribel Castillo-Caicedo (2011), "Deserción y graduación estudiantil universitaria: una aplicación de los modelos de supervivencia", en Revista Iberoamericana de Educación Superior (ries), México, unam-iisue/Universia, Vol. III, núm. 6, pp. 31-57,

<https://ries.universia.net/article/viewFile/72/130> [14 de Diciembre de 2014].

ANEXOS

ANEXO A

Listado de estudiantes seleccionados en la muestra para AEA

44096	54169	64202	200721003	200910573
44098	54175	69065	200720587	200910860
44092	54186	69067	200720642	200911287
44093	54187	69076	200720397	200910544
44082	54190	69077	200720639	200911049
44090	54194	69078	200720252	200910237
44101	54203	69080	200720327	200911843
49167	54205	69084	200720775	200910009
49174	59162	69089	200810587	200911818
49179	59172	69091	200811637	200911659
49184	59161	69092	200810140	200910689
49260	59144	69094	200810699	200910769
59180	59151	69095	200810181	200910873
49194	59152	69096	200810214	200921306
49256	59169	69100	200811624	200921187
49202	59183	69102	200810297	200922205
49168	64161	200711404	200810629	200920504
49177	64163	200711263	200810592	200920762
49204	64177	200711237	200810589	200921093
54181	64178	200711521	200810607	200922503
54188	64185	200711613	200810174	200921331
54192	64194	200711028	200820580	200920623
54193	64195	200711016	200821035	200921369
54312	64196	200711584	200820480	200921473
54172	64197	200711599	200820084	200921424
54176	64174	200711152	200820451	200920023
54191	64183	200711695	200821476	200921277
54267	64184	200711431	200820501	200922372
54183	64191	200711378	200820620	200921388
54197	64198	200720266	200820062	200921336
54171	64200	200720582	200910592	200921146

ANEXO B

Listado de estudiantes seleccionados en la muestra para AI

44040	54165	64151	200720288	200820412
44047	54271	64152	200720316	200820496
44053	54296	64154	200720369	200820568
44055	59046	64156	200720403	200820571
44060	59049	64158	200720424	200820579
44065	59055	64159	200720622	200820584
44070	59057	69155	200720784	200820596
44077	59060	69161	200720821	200821531
44078	59062	69163	200720836	200821691
44079	59073	69165	200720861	200910013
44250	59075	69166	200720870	200910207
49209	59076	69172	200720903	200910462
49211	59080	69174	200721360	200910647
49212	59085	69175	200721890	200910819
49217	59088	69182	200721894	200910899
49219	59092	69189	200722072	200910942
49222	59256	69191	200810037	200910956
49224	59261	69193	200810042	200910984
49229	59275	69194	200810119	200911099
49235	59285	200711010	200810239	200912068
49236	59288	200711052	200810246	200920028
49238	59292	200711081	200810550	200920347
49244	64114	200711158	200810559	200921096
49265	64115	200711333	200810561	200921113
49269	64119	200711349	200810602	200921198
54116	64120	200711386	200810625	200921250
54122	64129	200711408	200810645	200921255
54123	64130	200711478	200810743	200921261
54124	64132	200711498	200820010	200921322
54129	64135	200711633	200820036	200921356
54144	64137	200712002	200820083	200921472
54150	64140	200712297	200820106	200921504
54154	64142	200712625	200820109	200921560
54158	64146	200720097	200820181	20092195
54160	64147	200720108	200820354	
54164	64150	200720113	200820377	

ANEXO C

Listado de estudiantes seleccionados en la muestra para ATH

44121	54230	64239	200711470	200810001
44122	54231	64241	200711533	200810071
44124	54237	64245	200711544	200810078
44131	54238	64248	200711550	200810113
44134	54239	64249	200711607	200810141
44139	54251	69103	200711645	200810212
44140	54279	69104	200712142	200810222
44248	54284	69107	200720011	200810269
44251	59099	69111	200720030	200810313
49128	59101	69112	200720040	200810424
49129	59104	69113	200720079	200810485
49130	59105	69119	200720099	200810503
49132	59108	69120	200720105	200810575
49136	59109	69126	200720565	200810581
49147	59118	69128	200720586	200810591
49148	59121	69132	200720600	200810727
49151	59128	69138	200720661	200810736
49153	59131	69139	200720703	200810738
49155	59253	69140	200720706	200810909
49156	59257	69141	200720747	200811112
49159	64213	69142	200720908	200812153
49248	64217	69147	200720923	200820026
49254	64218	69149	200720952	200820127
49272	64220	200711053	200720979	200820137
54214	64224	200711202	200720984	200820197
54215	64226	200711213	200721001	200820266
54218	64232	200711270	200721005	200820279
54220	64235	200711423	200721083	200820315
54225	64236	200711442	200721534	
54228	64237	200711461	200721873	
200820394	200820884	200911673	200921288	20092247
200820490	200821097	200911799	200921324	
200820585	200910061	200911875	200921367	
200820601	200910299	200911876	200921438	
200820672	200910364	200911879	200921524	
200820689	200910797	200921240	200921754	

#SELECCIÓN DE LA MUESTRA#

Selección de la muestra: esta se hizo con el paquete de **R** llamado: "TeachingSampling", para cada programa como sigue:

1. Administración de Empresas Agropecuarias:

```
> AEA<-readXL ("D:/MIS DOCUMENTOS /EXCEL/ BD.xls ", header = T)
>Muestra1 <-S.SI (464,155)
>Muestra_AEA<-AEA[Muestra1]
```

2. Administración Turística y Hotelera:

```
> AI<-readXL ("D:/MIS DOCUMENTOS /EXCEL/ BD.xls ", header = T)
>Muestra2 <-S.SI (489,173)
>Muestra_AI<-AI[Muestra2]
```

3. Administración Industrial:

```
> ATH<-readXL ("D:/MIS DOCUMENTOS /EXCEL/ BD.xls ", header = T)
>Muestra3 <-S.SI (536, 178)
>Muestra_ATH<-ATH[Muestra3]
```

ANEXO D

MÉTODO DE SELECCIÓN DEL MODELO EN DESERCIÓN POR PROGRAMA

Éste se realiza HACIA ADELANTE, bajo el criterio AIC para los factores asociados a Deserción en AEA, se tiene que:

Start: AIC=434.63		
	Df	AIC
Género	1	434.00
Ingresos.Familiares	2	434.40
<none>		434.63
Tipo.de.Vivienda	1	435.00
Nivel.educativo.madre	3	435.13
Edad.de.ingreso.al.programa	1	436.17
Puntaje.estandarizado.ICFES	1	436.48
Trabajaba.cuando.presento.ICFE	1	436.55
No..De.hermanos	1	436.61

Surv(tiempos.de.Vida, Censurados) ~ Género		
Step 1: AIC=434		
	Df	AIC
Ingresos.Familiares	2	433.52
<none>		434.00
Tipo.de.Vivienda	1	434.50
Nivel.educativo.madre	3	435.33
Puntaje.estandarizado.ICFES	1	435.76
Edad.de.ingreso.al.programa	1	435.98
No..De.hermanos	1	435.99
Trabajaba.cuando.presento.ICFES	1	436.00

Surv(tiempos.de.Vida, Censurados) ~ Género + Ingresos.Familiares		
Step 2: AIC=433.52		
	Df	AIC
<none>		432.72
Tipo.de.Vivienda	1	432.71
Nivel.educativo.madre	3	433.90
X.Materias.perdidas.	1	434.48
Edad.de.ingreso.al.programa	1	434.64
Trabajaba.cuando.presento.ICFES	1	434.65
Puntaje.estandarizado.ICFES	1	434.71
No..De.hermanos	1	434.72

De modo que, el modelo óptimo queda determinado por el género e ingresos familiares. Sin embargo desde el marco conceptual del problema se decide considerar en el modelo la edad de Ingreso al programa. Giovagnoly (2002).

Luego para el programa de AI, se tiene que:

Start: AIC=649.05		
Surv(tiempos.de.Vida, Censurados) ~ 1		
	Df	AIC
Género	1	639.65
<none>		649.05
No..De.hermanos	1	650.39
Edad.de.ingreso.al.programa	1	650.47
Trabajaba.cuando.presento.ICFES	1	650.61
Puntaje.estandarizado.ICFES	1	650.91
Tipo.de.Vivienda	1	650.98
Ingresos.Familiares	2	652.22
Nivel.educativo.madre	3	654.36

Step 1: AIC=639.65		
Surv(tiempos.de.Vida,Censurados)-Género		
	Df	AIC
<none>		639.65
Trabajaba.cuando.presento.ICFES	1	640.41
Edad.de.ingreso.al.programa	1	640.90
No..De.hermanos	1	640.98
Puntaje.estandarizado.ICFES	1	641.08
Tipo.de.Vivienda	1	641.63
Ingresos.Familiares	2	643.32
Nivel.educativo.madre	3	645.14

Step 2: AIC=638.24		
Surv(tiempos.de.Vida,Censurados)-Género		
	Df	AIC
<none>		638.24
Puntaje.estandarizado.ICFES	1	639.19
Edad.de.ingreso.al.programa	1	639.31
Trabajaba.cuando.presento.ICFES	1	639.69
No..De.hermanos	1	639.77
Tipo.de.Vivienda	1	640.16
Ingresos.Familiares	2	642.10
Nivel.educativo.madre	3	643.69

Con esto el modelo óptimo queda determinado por el género. Sin embargo desde el marco conceptual del problema se decide considerar en el modelo la edad de Ingreso al programa. Giovagnoly (2002).

Finalmente para los factores asociados a la Deserción en ATH, según el método de selección del modelo hacia adelante y bajo el criterio AIC, se muestra en detalle como sigue:

Start:AIC=306.87		
Surv(tiempos.de.Vida,Censurados)=1		
	Df	AIC
Género	1	303.10
Ingresos.Familiares	2	304.26
Puntaje.estandarizado.ICFES	1	304.77
<none>		306.87
Edad.de.ingreso.al.programa	1	307.54
Trabajaba.cuando.presentoICFES	1	308.18
Tipo.de.Vivienda	1	308.56
No..De.hermanos	1	308.76
Nivel.educativo.madre	3	310.76

Step 1:AIC=303.09		
Surv(tiempos.de.Vida,Censurados)~Género		
	Df	AIC
Puntaje..ICFES	1	300.56
Ingresos.Familiares	2	301.02
Edad.de.ingreso.al.programa	1	302.77
<none>		303.10
Trabajaba.cuando.presentoICFES	1	304.68
Tipo.de.Vivienda	1	304.91
No..De.hermanos	1	305.03
Nivel.educativo.madre	3	306.71

Step 2:AIC=300.56		
Surv(tiempos.de.Vida,Censurados)~Género+ Puntaje.estandarizado.ICFES		
	Df	AIC
Ingresos.Familiares	2	298.57
Edad.de.ingreso.al.programa	1	299.25
<none>		300.56
Tipo.de.Vivienda	1	302.37
Trabajaba.cuando.presentoICFES	1	302.43
No..De.hermanos	1	302.44
Nivel.educativo.madre	3	304.38

Step 3:AIC=298.57		
Surv(tiempos.de.Vida,Censurados)~Género+ Puntaje.estandarizado.ICFES+ Ingresos.Familiares		
	Df	AIC
Edad.de.ingreso.al.programa	1	297.61
<none>		298.57
Trabajaba.cuando.presentoICFES	1	300.27
No..De.hermanos	1	300.34
Tipo.de.Vivienda	1	300.46
Nivel.educativo.madre	1	304.42

Step 5: AIC=297.61		
Surv(tiempos.de.Vida,Censurados)~Género+Puntaje.estandarizado.ICFES+Ingresos.Familiares +Edad.de.ingreso.al.programa		
	Df	AIC
<none>		297.61
No..De.hermanos	1	299.35
Estado.de.empleo.cuando.presento.ICFES	1	299.37
Tipo.de.Vivienda	1	299.60
Nivel.educativo.madre	3	303.52

El modelo óptimo queda determinado por el género, puntaje del ICFES estandarizado, edad de ingreso al programa e ingresos familiares de Nivel Medio y Bajo.

ANEXO E

***** Sintaxis en "R"*****

ANALISIS BIVARIADO#

Se laboraron tablas de contingencia para el programa AEA, para la variable dicotómica (Desertó) contrastado algunas cualitativas, posteriormente con la respectiva prueba chi-cuadrado, además se muestra la programación para la elaboración de los histogramas:

#Estado de empleo cuando presentó el ICFES vs Desertó#

```
>empleo<- xtabs(~Estado.de.empleo.cuando.presento.ICFES+Desertó, data=AEA)
> .Test <- chisq.test(empleo, correct=FALSE)
>barplot(empleo,main="Desertó vs Estado.de.empleo.cuando.presento.ICFES ",
xlab="Desertó",ylab="Número de Estudiantes",legend=c("Trabajaba","No
trabajaba"), beside=TRUE, col = topo.colors (2))
```

#Tipo de Vivienda vs Desertó#

```
>tipvi <- xtabs(~Tipo.de.Vivienda+Desertó, data=AEA)
>.Test <- chisq.test(tipvi, correct=FALSE)
> barplot(tipvi,main= "Tipo.de.Vivienda vs Desertó" xlab="Desertó", ylab="
Tipo.de.Vivienda", legend=c("Posee","Carece"), beside=TRUE, col = heat.colors
(2))
```

Género vs Desertó#

```
>gen <- xtabs(~Género+Desertó, data=AEA)
>.Test <- chisq.test(gen, correct=FALSE)
>barplot(genero,main="Estado del Estudiante vs Género", xlab="Desertó",
ylab="Número de Estudiantes",legend=c("Hombre","Mujer"), beside=TRUE,
col = topo.colors (2))
```

#Ingresos Familiares vs Desertó#

```
>ingr <- xtabs(~Ingresos.Familiares+Desertó, data=AEA)
>.Test <- chisq.test(ingr , correct=FALSE)
```

```
>barplot(ingr,width=1,      main="Ingresos      Familiares      vs      Desetó",
xlab="Desertó",ylab="Número de Estudiantes",legend=c("[0,1)", "Alto", "Medio",
"Bajo" ,"SC"), beside=TRUE, col=rainbow(7))
```

#Nivel educativo de la madre vs Desertó#

```
>nem<- xtabs(~Nivel educativo de la madre vs Desertó, data=AEA)
>.Test(-cisq.test(nem,correct=FALSE))
>barplot(nem,width=1,  main="Nivel Educativo de la Madre vs Desetó",
xlab="Desertó",ylab="Número de estudiantes",legend=c("Primaria","Secundaria",
"Tecnológica","SC","Universitaria"), beside=TRUE, col=rainbow(5))
```

#Posteriormente se efectuaron pruebas de correlación para la variable Tiempo hasta la deserción vs variables cuantitativas continuas#

```
> cor.test(X.materias.perdidas,      tiempos.de.vida,      alternative="two.sided",
method="pearson"), data=AEA)

> cor.test(Número.de.materias.repetidas, tiempos.de.vida, alternative="two.sided",
method="pearson"), data=AEA)

> cor.test(Edad.de.ingreso.al.programa, tiempos.de.Vida,  alternative="two.sided",
method="pearson" , data=AEA)

> cor.test(Puntaje.estandarizado.ICFES, tiempos.de.Vida, alternative="two.sided",
method="pearson" , data=AEA)

> cor.test(No..De.hermanos, tiempos.de.Vida, alternative="two.sided",
method="pearson", data=AEA)
```

###Pruebas de igualdad de las funciones de sobrevivencia para los programas de Administración###

****Estimación de la función de Sobrevida por Kaplan-Meier para cada programa de Administración****

```
> .Survfit <- survfit(Surv(Tiempos.de.vida, Censurados, type="right") ~ Programa_Aca,
  conf.type="log", conf.int=0.95, type="kaplan-meier", error="greenwood",
  data=TODO_ADMIN)
  legend(("bottomleft", legend=c("Admon. de Empresas Agropecuarias.", "Admon.
  Industrial", "Admon. Turística y Hotelera"), title=" Función de Sobrevida por
  programa académico ", col=1:3, lty=1:3, bty="n")
> plot(.Survfit, col=1:3, lty=1:3, conf.int=FALSE, mark.time=TRUE,
  xlab="Semestres", ylab="Probabilidad")
```

##Test de Long-Rank para Igualdad de las Curvas por Género###

```
> survdiff(Surv(Tiempos.de.vida, Censurados) ~ Programa_Aca, rho=0, data=
  TODO_ADMIN)
```

####Estimación de la función de Sobrevida por Kaplan-Meier para las Administraciones por Género####

```
> .Survfit <- survfit(Surv(Tiempos.de.vida, Censurados, type="right") ~ GENERO,
  conf.type="log", conf.int=0.95, type="kaplan-meier", error="greenwood",
  data=TODO_ADMIN)
  legend(("bottomleft", legend=c("Mujer", "Hombre"), title="Género", col=1:2, lty=1:2,
  bty="n"), xlab="Semestres", ylab="Probabilidad")
> plot(.Survfit, col=1:2, lty=1:2, conf.int=FALSE, mark.time=TRUE)
```

##Test de Long-Rank para Igualdad de las Curvas por Género###

```
> survdiff(Surv(Tiempos.de.vida, Censurados) ~ GENERO, rho=0, data= TODO_ADMIN)
```

ANEXO F

*****DIAGRAMAS DE LEXIS POR PROGRAMAS*****

***** Se recomienda cargar el paquete "Epi" *****

---##Diagrama de Lexis para AEA##---

```
>LL <- Lexis.diagram( age=c(-1,13), date=c(2003,2015), entry.age=
"Edad.de.ingreso.al.programa", exit.age= "EdadEgreso", birth.date="Cohorte",
fail=(Censurados), lwd.life=1, cex.fail=0.8, col.fail=c("blue","red"), alab="Tiempo de
deserción", dlab="Cohorte", int=c(1,1),data=AEA)
abline( v=c(2004,2015), col="blue")
## Identificar las personas de estudio##
>text( LL$exit.date, LL$exit.age, paste(1:nrow(LL)), col="Black", font=10, adj=c(0,1))
```

---##Diagrama de Lexis para AI##---

```
>LL <- Lexis.diagram( age=c(-1,15), date=c(2003,2015), entry.age=
"Edad.de.ingreso.al.programa", exit.age= "EdadEgreso", birth.date="Cohorte",
fail=(Censurados), lwd.life=1, cex.fail=0.8, col.fail=c("blue","red"), alab="Tiempo de
deserción",dlab="Cohorte", int=c(1,1),data=AI)
abline( v=c(2004,2015), col="blue")
---## Identificar las personas de estudio##--
>text( LL$exit.date, LL$exit.age, paste(1:nrow(LL)), col="Black", font=10,
adj=c(0,1))
```

---##Diagrama de Lexis para ATH##---

```
>LL <- Lexis.diagram( age=c(-1,15), date=c(2003,2015), entry.age=
"Edad.de.ingreso.al.programa", exit.age= "EdadEgreso", birth.date="Cohorte",
fail=(Censurados), lwd.life=1, cex.fail=0.8, col.fail=c("blue","red"), alab="Tiempo de
deserción",dlab="Cohorte", int=c(1,1),data=ATH)
abline( v=c(2004,2015), col="blue")
---## Identificar las personas de estudio##--
>text( LL$exit.date, LL$exit.age, paste(1:nrow(LL)), col="Black", font=10, adj=c(0,1))
```

```

***** SINTAXIS EN R *****
****ANÁLISIS PARA ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS****
#Definiendo a los Sobrevivientes#
attr(AEA_Surv, "time1") <- "PERMA_CEN"
attr(AEA_Surv, "event") <- "Censurados"

#Estimado la función de sobrevida por Kaplan- Meier#

>.Survfit <- survfit(Surv(PERMA_CEN, Censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=AEA_Surv)
>summary(.Survfit)

##Graficar la función de sobrevida con algunos ajustes gráficos#
>plot(.Survfit, conf.int=FALSE, mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Sobrevida para AEA",lwd=2,col="blue")
    axis(1, seq(0,13,1))
    axis(2, seq(0,1,.1))

##Graficar la función de Riesgo Acumulado##
>plot(.Survfit, conf.int=FALSE,
fun="cumhaz",mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Riesgo acumulado para AEA",lwd=2,col="red")
    axis(1, seq(0,13,1))
    axis(2, seq(0,1,.1))
    abline( v =1,lty=3,col = "blue")
    abline( v =5,lty=3,col = "blue")

#MODELO DE DESERCIÓN CON TODAS LAS COVARIABLES#
> CoxModel.AEA <- coxph(Surv(tiempos.de.Vida,Censurados) ~
Edad.de.ingreso.al.programa + Estado.de.empleo.cuando.presento.ICFES + Género +
Ingresos.Familiares + Nivel.educativo.madre + No..De.hermanos +
Puntaje.estandarizado.ICFES + Tipo.de.Vivienda, method="efron", data=AEA_Surv)
>summary(CoxModel.AEA)

#SELECCIÓN EL MODELO#
> stepwise(CoxModel.AEA, direction='forward', criterion='AIC')
#MODELO AJUSATADO #
>CoxModel.AEA_DEF <- coxph(Surv(tiempos.de.Vida,Censurados) ~
Edad.de.ingresov.al.programa + Género + Ingresos.Familiares , method="efron",
data=AEA_Surv)
> summary(CoxModel.AEA_DEF)

```

#FUNCIÓN DE SOBREVIVENCIA PARA EL MODELO DE COX#

```
>.Survfit <- survfit(Surv(CoxModel.AEA_DEF)
>summary(.Survfit)
>plot(.Survfit, conf.int=FALSE, mark.time=TRUE,xlab="Tiempo
(Semestres)",ylab="Probabilidad ",main="Función de Riesgo Acumulada para
AEA",col="purple")
  abline(v=1,col="blue", lty=2)
  abline(v=5,col="blue", lty=2)
  axis(1, seq(0,13,1))
  axis(2, seq(0,1,.1))
```

#FUNCIÓN DE RIESGO ACUMULADA PARA EL MODELO DE COX#

```
>plot(.Survfit, fun="cumhaz",conf.int=FALSE, mark.time=TRUE,xlab="Tiempo
(Semestres)",ylab="Probabilidad ",main="Función de Riesgo Acumulada para
AEA",col="orange")
  abline(v=1,col="blue", lty=2)
  abline(v=5,col="blue", lty=2)
  axis(1, seq(0,13,1))
  axis(2, seq(0,1,.1))
```

###EVALUACIÓN DEL MODELO###

#Test de riesgos proporcionales#

```
> .CoxZPH <- cox.zph(CoxModel.AEA_DEF)
```

#Residuos de Cox-Snell#

```
>estado<-AEA_Surv$Censurados
>mresi<-residuals(CoxModel.AEA_DEF, type="martingale")
>csresi<-estado-mresi
>hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")
>plot(hazard.csresi$time ,-log(hazard.csresi$surv), xlab="residuos de Cox-Snell",
ylab="Riesgo acumulado",lty = 1:1, main="Representación de los residuos de Cox-Snell")
  lines(c(0,5),c(0,5))
```

#MODELO DE REGRESIÓN DE COX PARA GRADUACIÓN EN AEA

```
>Grad.AEA <- coxph(Surv(tiempo_Grad, Cens_Grad)~ Edad.de.ingreso.al.programa
+Estado.de.empleo.cuando.presento.ICFES +Género + Ingresos.Familiares
+Nivel.educativo.madre No..De.hermanos +Número.de.materias.repetidas
+Puntaje.estandarizado.ICFES + Tipo.de.Vivienda + X.Materias.perdidas., data=AEA_G)
```

```
summary(Grad.AEA)
```

```
#Selección del modelo#
```

```
>stepwise(Grad.AEA, direction='forward', criterion='AIC')
```

```
#Modelo Ajustado #
```

```
> Grad.AEA_AJUST <- coxph(Surv(tiempo_Grad, Cens_Grad)~ Género +  
Ingresos.Familiares + Tipo.de.Vivienda + úmero.de.materias.repetidas, data=AEA_G)  
>summary(Grad.AEA_AJUST)
```

```
##Función de sobrevida para el modelo ajustado ##
```

```
>.Survfit <- survfit(Surv(Grad.AEA_AJUST)  
>plot(.Survfit, ,conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de sobrevida para Coxmodel  
AEA",col="blue")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

```
##Función de Riesgo Acumulada para el modelo Ajustado##
```

```
>plot(.Survfit, ,conf.int=FALSE,fun= "cumhaz", mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de sobrevida para Coxmodel  
AEA",col="orange")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))  
abline( v =1,lty=3,col = "blue")  
abline( v =5,lty=3,col = "blue")
```

```
###EVALUACIÓN DEL MODELO###
```

```
#Test de riesgos proporcionales#
```

```
> .CoxZPH <- cox.zph(Grad.AEA_AJUST)
```

```
#Residuos de Cox-Snell#
```

```
>estado<- AEA_G$ Cens_Grad  
>mresi<-residuals(Grad.AEA_AJUST, type="martingale")  
>csresi<-estado-mresi  
>hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")  
>plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab="residuos de Cox-Snell",  
ylab="Riesgo acumulado",lty = 1:1, main="Representación de los residuos de Cox-Snell")  
lines(c(0,5),c(0,5))
```

***** ANÁLISIS DE DATOS PARA EL PROGRAMA DE ADMINISTRACIÓN*****
 *****INDUSTRIAL*****

-----Sintaxis en **R**-----

#Definiendo a los Sobrevivientes#

```
attr(AI_Surv, "time1") <- " tiempos.de.Vida "
attr(AI_Surv, "event") <- "Censurados"
```

#Estimado la función de sobrevida por Kaplan- Meier#

```
>.Survfit <- survfit(Surv(tiempos.de.Vida, Censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=AI_Surv)
>summary(.Survfit)
```

##Graficar la función de sobrevida con algunos ajustes gráficos#

```
>plot(.Survfit, conf.int=FALSE, mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Sobrevida para AI",lwd=2,col="blue")
  axis(1, seq(0,13,1))
  axis(2, seq(0,1,.1))
```

##Graficar la función de Riesgo Acumulado##

```
>plot(.Survfit, conf.int=FALSE,
fun="cumhaz" mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Riesgo acumulado para AI",lwd=2,col="red")
  axis(1, seq(0,13,1))
  axis(2, seq(0,1,.1))
  abline( v =1,lty=3,col = "blue")
  abline( v =5,lty=3,col = "blue")
```

#MODELO DE DESERCIÓN CON TODAS LAS COVARIABLES#

```
> CoxModel.AI <- coxph(Surv(tiempos.de.Vida,Censurados) ~ Edad.de.ingreso.al.programa
+ Estado.de.empleo.cuando.presento.ICFES + Género + Ingresos.Familiares +
Nivel.educativo.madre + No..De.hermanos + Puntaje.estandarizado.ICFES +
Tipo.de.Vivienda, method="efron", data=AI_Surv)
>summary(CoxModel.AI)
```

#SELECCIÓN EL MODELO#

```
> stepwise(CoxModel.AI, direction='forward', criterion='AIC')
```


#MODELO AJUSATADO

```
>CoxModel.AI_DEF <- coxph(Surv(tiempos.de.Vida,Censurados) ~  
Edad.de.ingresov.al.programa , method="efron", data=AI_Surv)  
> summary(CoxModel.AI_DEF)
```

#FUNCIÓN DE SOBREVIDA PARA EL MODELO DE COX#

```
>.Survfit<-(survfit(CoxModel.AI_DEF))  
>summary(.Survfit)  
>plot(.Survfit, ,conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de Sobrevida CoxModel para  
Administración Industrial", col="red")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

#FUNCIÓN DE RIESGO ACUMULADA PARA EL MODELO DE COX#

```
>plot(.Survfit, fun="cumhaz",conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de Riesgo Acumulada para  
AI",col="red")  
abline(v=1,col="blue", lty=2)  
abline(v=5,col="blue", lty=2)  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

###EVALUACIÓN DEL MODELO###

#Test de riesgos proporcionales#

```
> .CoxZPH <- cox.zph(CoxModel.AI_DEF)
```

#Residuos de Cox-Snell#

```
>estado<-AI_Surv$Censurados  
>mresi<-residuals(CoxModel.AI_DEF, type="martingale")  
>csresi<-estado-mresi  
>hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")  
>plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab="residuos de Cox-Snell",  
ylab="Riesgo acumulado",lty = 1:1, main="Representación de los residuos de Cox-Snell")  
lines(c(0,5),c(0,5))
```

*****MODELO DE REGRESIÓN DE COX PARA GRADUACIÓN EN AI*****

```
> Grad.AI <- coxph(Surv(tiempo_Grad, Cens_Grad)~ Edad.de.ingreso.al.programa  
+ Estado.de.empleo.cuando.presento.ICFES + Género + Ingresos.Familiares  
+ Nivel.educativo.madre + No..De.hermanos + Número.de.materias.repetidas  
+ Puntaje.estandarizado.ICFES + Tipo.de.Vivienda + X.Materias.perdidas., data=AI_G)  
> summary(Grad.AI)
```

#Selección del modelo#

```
> stepwise(Grad.AI, direction='forward', criterion='AIC')
```

#Modelo Ajustado #

```
> Grad.AI_AJUST <- coxph(Surv(tiempo_Grad, Cens_Grad)~ Género +  
Ingresos.Familiares + Tipo.de.Vivienda + número.de.materias.repetidas, data=AEA_G)  
> summary(Grad.AI_AJUST)
```

##Función de sobrevida para el modelo ajustado ##

```
> .Survfit <- survfit(Surv(Grad.AI_AJUST)  
> plot(.Survfit, ,conf.int=FALSE,  
mark.time=TRUE,xlab="Tiempo(Semestres)",ylab="Probabilidad ",main="Función de  
sobrevida para CoxModel AI",col="red")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

##Función de Riesgo Acumulada para el modelo Ajustado##

```
> plot(.Survfit, ,conf.int=FALSE,fun= "cumhaz", mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de sobrevida para Coxmodel  
AEA",col="green")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

###EVALUACIÓN DEL MODELO###

#Test de riesgos proporcionales#

```
> .CoxZPH <- cox.zph(Grad.AI_AJUST)
```

#Residuos de Cox-Snell#

```
> estado <- AI_G$ Cens_Grad  
> mresi <- residuals(Grad.AI_AJUST, type="martingale")  
> csresi <- estado-mresi  
> hazard.csresi <- survfit(Surv(csresi,estado)~1,type="fleming-harrington")
```

```
>plot(hazard.csresi$time, -log(hazard.csresi$surv), xlab="residuos de Cox-Snell",
ylab="Riesgo acumulado", lty = 1:1, main="Representación de los residuos de Cox-Snell")
lines(c(0,5),c(0,5))
```

```
***** ANÁLISIS DE DATOS PARA EL PROGRAMA DE ADMINISTRACIÓN*****
*****TURÍSTICA Y HOTELERA*****
```

```
-----Sintaxis en R-----
```

```
#Definiendo a los Sobrevivientes#
```

```
attr(ATH_Surv, "time1") <- " tiempos.de.Vida "
attr(ATH_Surv, "event") <- "Censurados"
```

```
#Estimado la función de sobrevida por Kaplan- Meier#
```

```
>.Survfit <- survfit(Surv(tiempos.de.Vida, Censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=ATH_Surv)
>summary(.Survfit)
```

```
##Graficar la función de sobrevida con algunos ajustes gráficos#
```

```
>plot(.Survfit, conf.int=FALSE, mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Sobrevivencia CoxModel Grad",lwd=2,col="blue")
axis(1, seq(0,13,1))
axis(2, seq(0,1,.1))
```

```
##Graficar la función de Riesgo Acumulado##
```

```
>plot(.Survfit, conf.int=FALSE,
fun="cumhaz",mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",
main="Función de Riesgo acumulado para ATH",lwd=2,col="red")
axis(1, seq(0,13,1))
axis(2, seq(0,1,.1))
abline(v=1,lty=3,col="blue")
abline(v=5,lty=3,col="blue")
```

```
#MODELO DE DESERCIÓN CON TODAS LAS COVARIABLES#
```

```
> CoxModel.ATH <- coxph(Surv(tiempos.de.Vida,Censurados) ~
Edad.de.ingreso.al.programa + Estado.de.empleo.cuando.presento.ICFES + Género +
Ingresos.Familiares + Nivel.educativo.madre + No..De.hermanos +
Puntaje.estandarizado.ICFES + Tipo.de.Vivienda, method="efron", data=ATH_Surv)
>summary(CoxModel.ATH)
```

#SELECCIÓN EL MODELO#

```
> stepwise(CoxModel.ATH, direction='forward', criterion='AIC')
```

#MODELO AJUSATADO

```
> CoxModel.ATH_DEF <- coxph(Surv(tiempos.de.Vida, Censurados) ~  
Edad.de.ingresov.al.programa + Género + Ingresos.Familiares , method="efron",  
data=ATH_Surv)  
> summary(CoxModel.ATH_DEF)
```

#FUNCIÓN DE SOBREVIDA PARA EL MODELO DE COX#

```
>.Survfit <- survfit(Surv(CoxModel.AEA_DEF)  
>summary(.Survfit)  
>plot(.Survfit, ,conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de Supervida para ATH",col="blue")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

#FUNCIÓN DE RIESGO ACUMULADA PARA EL MODELO DE COX#

```
>plot(.Survfit, fun="cumhaz",conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de Riesgo Acumulada para  
ATH",col="red")  
abline(v=1,col="blue", lty=2)  
abline(v=5,col="blue", lty=2)  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

###EVALUACIÓN DEL MODELO###

#Test de riesgos proporcionales#

```
> .CoxZPH <- cox.zph(CoxModel.ATH_DEF)
```

#Residuos de Cox-Snell#

```
>estado<-ATH_Surv$Censurados  
>mresi<-residuals(CoxModel.ATH_DEF, type="martingale")  
>csresi<-estado-mresi  
>hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")  
>plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab="residuos de Cox-Snell",  
ylab="Riesgo acumulado",lty = 1:1, main="Representación de los residuos de Cox-Snell")  
lines(c(0,5),c(0,5))
```

*****MODELO DE REGRESIÓN DE COX PARA GRADUACIÓN EN ATH CON
TODAS LAS COVARIABLES*****

```
>Grad.ATH <- coxph(Surv(tiempo_Grad, Cens_Grad)~ Edad.de.ingreso.al.programa  
+Estado.de.empleo.cuando.presento.ICFES +Género + Ingresos.Familiares  
+Nivel.educativo.madre No..De.hermanos +Número.de.materias.repetidas  
+Puntaje.estandarizado.ICFES + Tipo.de.Vivienda + X.Materias.perdidas., data=ATH_G)  
summary(Grad.ATH)
```

#Selección del modelo#

```
>stepwise(Grad.ATH, direction='forward', criterion='AIC')
```

#Modelo Ajustado #

```
> Grad.ATH_AJUST <- coxph(Surv(tiempo_Grad, Cens_Grad)~ +  
Edad.de.ingreso.al.programa +Número.de.materias.repetidas, data=ATH_G)  
>summary(Grad.ATH_AJUST)
```

##Función de sobrevida para el modelo ajustado ##

```
>.Survfit <- survfit(Surv(Grad.ATH_AJUST)  
>plot(.Survfit, ,conf.int=FALSE, mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de sobrevida para Coxmodel  
ATH",col="blue")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))
```

##Función de Riesgo Acumulada para el modelo Ajustado##

```
>plot(.Survfit, ,conf.int=FALSE,fun= "cumhaz", mark.time=TRUE,xlab="Tiempo  
(Semestres)",ylab="Probabilidad ",main="Función de sobrevida para Coxmodel  
ATH",col="orange")  
axis(1, seq(0,13,1))  
axis(2, seq(0,1,.1))  
abline( v =1,lty=3,col = "blue")  
abline( v =5,lty=3,col = "blue")
```

```

####EVALUACIÓN DEL MODELO####
#Test de riesgos proporcionales#
> .CoxZPH <- cox.zph(Grad.ATH_AJUST)
#Residuos de Cox-Snell#
>estado<- ATH_G$ Cens_Grad
>mresi<-residuals(Grad.ATH_AJUST, type="martingale")
>csresi<-estado-mresi
>hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")
>plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab="residuos de Cox-Snell",
ylab="Riesgo acumulado", lty = 1:1, main="Representación de los residuos de Cox-Snell")
lines(c(0,5),c(0,5))

```

ANEXO G
Pronóstico para la cohorte I de 2015

ADMINISTRACIÓN INDUSTRIAL				
Género	Edad.Ing	PIC	RR respecto al estudiante base	Riesgo
M	21	0,73227	2,079796391	+
F	20	-0,05195	0,949376334	-
M	17	0,94007	2,560160623	+
M	17	0,94007	2,560160623	+
F	16	0,15585	1,168650892	+
M	21	0,73227	2,079796391	+
M	19	0,83617	2,307512259	+
M	18	0,88812	2,430555908	+
F	20	-0,05195	0,949376334	-
F	17	0,1039	1,1094895	+
M	17	0,94007	2,560160623	+
M	18	0,88812	2,430555908	+
F	18	0,05195	1,053323075	+
F	17	0,1039	1,1094895	+
M	18	0,88812	2,430555908	+
M	23	0,62837	1,874552567	+
M	20	0,78422	2,19069753	+
F	17	0,1039	1,1094895	+
F	17	0,1039	1,1094895	+
F	17	0,1039	1,1094895	+
F	19	0	1	
M	23	0,62837	1,874552567	+
F	18	0,05195	1,053323075	+
F	20	-0,05195	0,949376334	-
F	19	0	1	
F	18	0,05195	1,053323075	+
F	18	0,05195	1,053323075	+
M	17	0,94007	2,560160623	+
F	17	0,1039	1,1094895	+
F	17	0,1039	1,1094895	+
F	18	0,05195	1,053323075	+
F	20	-0,05195	0,949376334	-
M	20	0,78422	2,19069753	+
F	19	0	1	
F	16	0,15585	1,168650892	+

M	18	0,88812	2,430555908	+
F	17	0,1039	1,1094895	+
M	19	0,83617	2,307512259	+
F	16	0,15585	1,168650892	+
F	20	-0,05195	0,949376334	-
F	17	0,1039	1,1094895	+
F	23	-0,2078	0,812369494	-
F	18	0,05195	1,053323075	+
F	16	0,15585	1,168650892	+
F	16	0,15585	1,168650892	+

En donde el signo (+) indica un mayor riesgo de deserción para los estudiantes que ingresan al programa de AI, especialmente en ser hombre y registrar una edad de 20 años respecto al estudiante base mencionado en el apartado 5.2.4.

A continuación se describe en la tabla para el programa de ATH:

ADMINISTRACIÓN TURISTICA Y HOTELERA						
Género	Punt.ICFES	Ingresos	Edad	PIC	RR respecto al estudiante base	Riesgo
M	85	Bajo	19	0,322256	1,38023807	+
M	96	Bajo	23	0,803656	2,2336924	+
M	58	Bajo	20	-0,110744	0,89516788	-
M	60	Bajo	18	-0,211944	0,80901	-
F	74	Bajo	19	1,375016	3,95514	+
F	60	Bajo	17	0,976216	2,65439299	+
F	80	Bajo	16	1,279016	3,59310237	+

En el caso del programa de ATH, se tiene un mayor riesgo de desertar cuando se registran ingresos familiares de nivel bajo, en ser hombre con un año menos de edad de ingreso.

Edad Ingres	Género	Ingresos Familiares	PIC	RR respecto al estudiante base	Riesgo
19	H	Bajo	0,33122809	1,39267741	+
18	H	Bajo	0,33237639	1,39427754	+
21	H	Bajo	0,32893149	1,38948266	+
19	H	Bajo	0,33122809	1,39267741	+
27	H	Medio	3,58130369	35,9203393	+
16	F	Bajo	-0,17442841	0,83993698	-
24	F	Bajo	-0,16639031	0,84671569	-
17	F	Bajo	-0,17098351	0,84283547	-
21	F	Bajo	-0,16868691	0,84477335	-
19	F	Bajo	-0,16639031	0,84671569	-
17	F	Bajo	-0,16639031	0,84671569	-
17	F	Bajo	-0,16868691	0,84477335	-
19	F	Bajo	-0,17328011	0,84090204	-
23	F	Bajo	-0,17442841	0,83993698	-
24	F	Bajo	-0,16639031	0,84671569	-

Para el programa de AEA, se tiene un mayor riesgo de desertar respecto de un estudiante típico, (mencionado en la sección 6.2.4) cuando se es hombre y se registra tener ingresos familiares de nivel medio.

ANEXO H

FORMATO DE REGISTRO DE DATOS

Teniendo en cuenta que el análisis efectuado en este proyecto indicó un pobre ajuste, es importante indicar algunas otras variables que podrían estar afectando el riesgo de deserción y la graduación. Lo anterior con el fin de que cada programa recoja la información pertinente a sus estudiantes y a futuro la utilice para construir modelos de sobrevida con una mayor bondad de ajuste.

Las variables que se proponen se dividen en 4 grupos, así:

I. Información Personal

- Semestre de ingreso: año del semestre en el que un estudiante se matriculó al programa, ejemplo: 2016-I, significa primer semestre del 2016.
- Código: Valor numérico asignado por registro en el momento de la matrícula.
- *Género*, los valores que toma son: Masculino (M), Femenino (F).
- *Edad*: valor numérico de la edad en la que se matriculó en el programa.
- Estado Civil: Soltero(a), Casado(a), Divorciado(a), Viudo(a)
- Si tiene hijos, indicar el número de estos.
- Si tiene hermanos, señalar el número de hermanos
- Posición entre hermanos: No tiene (0), primer lugar (1), segundo (2), así sucesivamente.
- Estado de la salud Física: Bueno, Malo, Regular.
- *Población Vulnerable*: Desplazado (DESPL), Afrodescendiente (AFRO), LGTBI, Ninguno.
- *Embarazos no planeados*: si ha experimentado tales, Si o No.
- *Calamidad y/o problema doméstico* los valores que toma: Si presentó algunos de los siguientes eventos, Muerte de los padres (MP), Muerte de un Hijo (MH), Ninguno.
- *Expectativas no satisfechas*: Percepciones de estudiante antes de ingresar al programa, posibles valores que puede tomar: no le gusta el programa(NGPR), no

era como se lo esperaba(NESP), es muy duro el cambio del colegio a la Universidad(CAMB_U)

- *Discapacidad*: si el estudiante manifiesta tenerla, posibles valores a tomar: Física, Sensorial: dentro de esta se encuentran (auditiva, visual), psíquica, intelectual o mental.

II. Información Socioeconómica

- Situación Laboral del estudiante al momento de ingresar a la universidad: valores que toma la variable: Si, No.
- ¿Cuánto es el nivel de ingresos?: Bajo (0, 1 SMMLV], Medio (1,3 SMMLV], Alto (3, en adelante], SMMLV: Salario Mínimo Mensual Legal Vigente.
- ¿Con quién vive? Madre y padre (1), Madre o padre (2), Otro familiar (3), ningún familiar (4)
- Si se tienen personas a cargo, indicar el número de estas.
- Nivel del Sisbén, indica el nivel de este, valores que toma (1,2,3)
- Clasificación del Estrato socioeconómico los valores que toma son (1,2,3,4,5)
- Nivel de Ingresos de los Padres, indica cual es el ingreso familiar de los padres, y los valores que puede tener: Bajo (0, 1 SMMLV], Medio (1,3 SMMLV], Alto (3, en adelante]
- Máximo Nivel de estudios alcanzado por el Padre, estos valores son: Ninguno, Primaria, Básica, Media, Técnico, Universitario-Superior.
- Máximo Nivel de estudios alcanzado por el Madre, estos valores son: Ninguno, Primaria, Básica, Media, Técnico, Universitario-Superior.
- Ocupación del padre: Jubilado, hogar, estudiante, busca empleo (Baja), trabajador independiente, empleado (Media), empresario, administrador, gerente, profesional independiente (Alta).
- Ocupación de la madre: Jubilada, hogar, estudiante, busca empleo (Baja), trabajador independiente, empleada (Media), empresaria, administradora, gerente, profesional independiente (Alta).

- Tipo de vivienda, Si es propia o No.
- Número de Hermanos, indica la cantidad de hermanos que tiene.
- Tipo de relación con los padres, la opinión del estudiante y su relación con sus progenitores, puede tomar los siguientes valores: Buena, mala, regular.


III. Variables Académicas

- *Tiempo de ingreso a la Universidad*: indica el número de semestres que el estudiante demoró en ingresar a la Universidad.
- Puntaje estandarizado de la prueba saber 11.
- Tipo de Colegio: indica el tipo de colegio de egreso del estudiante, tiene dos categorías Público o Privado.
- *Rendimiento académico durante el Colegio*: Cómo considera que fue este, puede ser: Excelente, Sobresaliente, Aceptable, Insuficiente.
- *¿Ha recibido algún tipo de orientación vocacional?* Indique Si o No.
- Énfasis del colegio de egreso de secundaria del estudiante: técnico agropecuario, comercial=1, académico=2, turístico=3, técnico=4.
- Orientación Vocacional previa recibida antes de ingresar a la universidad. Tiene dos categorías posibles: Sin O.V.=1 si no recibió ninguna orientación vocacional, O.V.=1 si recibió orientación.
- *¿Ha presentado matoneo (bullying)?* Indique Si o No.
- Empezar Universidad: Esta variable cuenta con cuatro categorías. E1: es la primera vez que el alumno comienza un estudio superior, COC: si además de la carrera iniciada, continua otra carrera, TOC: en caso en que el estudiante, al iniciar esta carrera, ya haya terminado previamente algún otro estudio superior. Y por último, AC: en caso que abandono una carrera anteriormente.

IV. Información Institucional

- ¿Ha recibido beca o financiamiento de la Universidad? Si, No.
- ¿El programa académico que cursa, cuenta con registro calificado? Si, No.
- ¿Considera que usted tiene una buena relación con los docentes? Si, No.
- ¿Considera que usted tiene una buena relación con sus compañeros? Si, No.
- ¿Ha iniciado otro programa académico en esta Institución u otra?, Indicar Si o No, además. Mencione ¿cuál?_____

VISTA DE LOS ENCABEZADOS DE LA HOJA DE EXCEL

														
INFORMACIÓN PERSONAL														
N°	SEMESTRE DE INGRESO	CÓDIGO	GÉNERO	EDAD	ESTADO CIVIL	N° DE HERMANOS	POSICIÓN ENTRE HERMANOS	N° HIJOS	ESTADO SALUD FISICA	POBLACIÓN VULNERABLE	EMBARAZOS NO PLANEADOS	CALAMIDAD Y/O PROBLEMAS DOMÉSTICOS	EXPECTATIVAS NO SASTISFECHAS	DISCAPACIDAD
1														
2														

FORMATO DE CAPTURA DE DATOS - VARIALES NO ESTATICAS

INFORMACIÓN SOCIOECONOMICA														
SITUACIÓN LABORAL	CUENTA PROPIA	NIVEL DE INGRESOS ESTUDIANTE	PERSONAS CON QUIEN VIVE	# PERSONAS A CARGO	NIVEL DEL SISBÉN	ESTRATO	NIVEL DE INGRESOS PADRES	NIVEL EDUCATIVO PADRE	NIVEL EDUCATIVO MADRE	OCUPACIÓN DEL PADRE	OCUPACIÓN DE LA MADRE	TIPO DE VIVIENDA	NUMERO DE HERMANOS	TIPO DE RELACIÓN CON PADRES



ACADÉMICAS										INSTITUCIONAL					
TIEMPO DE INGRESO A LA UNIV.	PUNTAJE P.SABER	TIPO COLEGIO	TIPO PROGRAMA ACADÉMICO	RENDIMIENTO ACADEMICO EN EL COLEGIO	ORIENTACION VOCACIONAL	Énfasis colegio de egreso	Orientación vocacional previa	Empezar Universidad	INSATISFACCIÓN CON EL PROGRAMA	NORMATIVAS ACADÉMICAS	BECAS/FINANCIAMIENTO	CALIDAD DEL PROGRAMA	RELACIÓN CON PROFESORES	RELACIÓN CON ESTUDIANTES	PROVIENE DE OTRO PROGRAMA